Success in rhythmic gymnastics competition: study of evaluation and performance variables in individual routines

Catarina Leandro

International Doctoral Thesis 2016

Supervisors: Professora Marta Bobo Arce

Professora Elena Sierra Palmeiro

Professora Maria de Lurdes Ávila-Carvalho.

Department of Physical Education and Sport

PhD in Sport, Physical Education and Healthy Leisure

UNIVERSIDADE DA CORUÑA

The undersigned thesis supervisors,

Marta Bobo Arce, Elena Sierra Palmeiro and Maria de Lurdes Ávila-Carvalho,

confirm that the doctoral thesis entitled

SUCCESS IN RHYTHMIC GYMNASTICS COMPETITION: STUDY OF EVALUATION AND PERFORMANCE VARIABLES IN INDIVIDUAL ROUTINES

produced by the candidate

CATARINA LEANDRO

fulfill the requirements for the International PhD degree
in the University of A Coruna.

_____ _____

Place and date

Professor Marta Bobo Arce

_____

Professor Elena Sierra Palmeiro

_____

Professor Maria de Lurdes Ávila-Carvalho

Dedication

Eunice Lebre

" Um médico que só sabe medicina, nem medicina sabe"

Dr. Abel salazar

Acknowledgements

Esta Dissertação resulta da colaboração atenta e empenhada de muitos dos meus, professores e colegas de trabalho, orientadores, entidades e personalidades da ginástica, familiares e amigos que diáriamente me foram dando incentivo, apoio, confiança, partilha de conhecimentos e sugestões. Aqui fica, com um profundo agradecimento a todos, os nomes daqueles que, de forma direta, estão dentro deste trabalho de investigação:

Orientadoras Marta Bobo, Elena Sierra Palmeiro e Maria de Lurdes Ávila-Carvalho, Federação Internacional de Ginástica, Juízes Internacionais, Federação de Ginástica de Portugal, Amigas Eunice Lebre, Isabel Antunes, Gabriela Salvador, Silvia Canelas e Ana Isabel Cardoso, Marido Emanuel Gonçalves, filhos Ricardo, Rui Pedro e Sofia, Prima Filomena Gonçalves, colegas de trabalho Raquel Santiago, Fátima Silva, Lúcia Bandeira, Ermelinda Azevedo, Maria João Félix André Barreiros, Paula Botelho Gomes, Carlos Lago...

Abstract

The main goal is the analysis of evaluation and performance variables in individual routines of world reference gymnasts at rhythmic gymnastics competition. In order to achieve this goal, 12165 technical elements and the answers of 162 international judges were analysed. The results indicate that most international judges have a background which enables a quality judgment. Furthermore, the Code of points prescribes evaluation criteria that may compromise the objective evaluation of the artistic faults and the difficulty elements: Mastery and Risk. The efficiency of these resources has not yet been reinforced, since we detected a disagreement between the 4 judges in 40% of the difficulty elements observed. The quality of the judgement, tends to be lower in gymnasts of medium level and the judges and coaches do not perceive the evaluation criteria of the difficulty elements identically. Through the results concerning the content analysis of the routines, we can see a common pattern in the composition of competition routines in the different gymnast levels and different apparatus routines. The technical elements: Rotations and Risk, together, represent more than 50% of the total value of compositions.

Keywords: Rhythmic Gymnastics, Body difficulty , Evaluation, Performance.

Resumen

El objetivo de este estudio es analizar las variables de evaluación y de rendimiento en ejercicios individuales de Gimnasia Rítmica. Para ello analizamos 12165 elementos y a 162 jueces internacionales. Los resultados indican que la mayoría de jueces internacionales ofrecen garantía en la calidad en su juicio y que el Código de Puntuación prescribe criterios de evaluación que pueden comprometer la objetividad de la evaluación, especialmente en las faltas artísticas, algunos elementos de dificultad corporal, elementos de Maestría y Riesgos. La eficacia de estos criterios no parece consolidada ya que hay desacuerdo en la evaluación de las 4 jueces, en el 40% de los elementos de dificultad observados. La calidad del juicio, tiende a ser inferior en gimnasta de nivel medio y jueces y entrenadoras no tienen la misma percepción relativa a los criterios de evaluación de los elementos que se valoran en el apartado de dificultad.

Los resultados referentes al análisis del contenido de los ejercicios verifican la existencia de una estructura común en las composiciones de los ejercicios de competición, tanto en los diferentes niveles de gimnastas, como en los diferentes aparatos. Los elementos de Rotaciones y Riesgo representan más del 50% del valor total de la composición.

Palabras-clave: Gimnasia Rítmica, Dificultad Corporal, Evaluación, Rendimiento.

Resumo

O obxectivo deste estudo é analizar as variables de avaliación e de rendemento e exercicios individuais de Ximnasia Rítmica. Analizáronse 12165 elementos e 162 xuíces internacionais. Os resultados indican que a maioría dos xuíces internacionais ofrecen garantía da calidade no seu xuizo e que o Código de Puntuación prescribe criterios de avaliación que poden comprometer a obxectividade da avaliación, especialmente en as faltas artísticas, algúns elementos de elementos de dificultade corporal, elementos de Mestría e Riscos. A eficacia destes criterios non parece consolidada porque hai diverxencia na avaliación das catro xuíces nun 40% dos elementos de dificultade observados. A calidade do xuízo tende a ser menor en gimnastas de nivel medio e xuíces e adestradores non teñen a mesma percepción sobre os criterios de avaliación dos elementos que son medidos no apartado de dificultade.

Os resultados relativos á análise do contido dos exercicios verifican a existencia dunha estrutura común nas composicións dos exercicios de competición tanto nos diferentes niveis de ximnastas, como nos diferentes aparatos. Os elementos de Rotación e Risco representan máis do 50% do valor total da composición.

Palabras chave: Ximnasia Rítmica, Dificultade Corporal, Avaliación, Rendemento.

Contents

Index of Figures

2. Empirical Studies: The judgment in Rhythmic Gymnastics high level competition routines

3. Empirical Studies: The content in Rhythmic Gymnastics high level competition routines.

Index of Tables

1. Introduction

2. Empirical Studies: The judgment in Rhythmic Gymnastics high level competition routines.

Study I

Study VI

Abbreviations

Ab	Absolute
CoP	Code of Points
CP	Código de Pontuación/ Código de Pontuação
Cα	Cronbach's alpha coefficient.
D	Difficulty
DER	Dynamic Elements With Rotation and Throw/ Elementos Dinámicos de rotación con lanzamiento
Dev.	Deviation
DDS	Departure Difficulty Score
E	Execution
e.g.	For example
et al.	And colleagues
FDS	Final Difficulty
FIG	Federação Internacional de Ginástica / International Federation of Gymnastics
GR	Gimnasia Rítmica / Ginástica Rítmica
ICC	Intra - Class Correlations
max.	Maximum
min.	Minimum
MixDif	Mixed Difficulties/ Dificultad mixta
n	number
R	Corrected item - total correlation
RF	Fouetté Rotations
RFF	Rotations on the flat foot or on other part of the body.
RG	Rhythmic Gymnastics
RPIV	Relevê Rotations (Pivot)
sd	Standard Deviation
SPSS	Statistical Package for the Social Sciences
vs.	Versus
WC	World Championship
1st	First
2nd	Second
3rd	Third

1. Introduction

1. Introduction

1.1. Contextualization

Rhythmic Gymnastics (RG) is one of the four Olympic disciplines inside gymnastics family. Each Olympic cycle the Code of Points (CoP) and General Regulations are updated in order to get a better judgement accuracy of the gymnasts during competitions, and improve the sport quality. Each CoP new configuration affects the technical content of elite gymnasts' individual routines.

RG is a sport with very complex motor skills and high level of artistic components (Ávila-Carvalho, 2012). The requirements in terms of difficulty and skills variety for gymnasts increase and their effort to accomplish all demands also. The complexity of the routines is raised and the gymnasts have to achieve always higher levels of conditional and coordinative skills. To judge all this work, constantly updated, the judges should have a deep knowledge of the sport.

The competitive process of the rhythmic gymnastics elite is therefore associated to a multiplicity of variables to which the sports agents must be able to work with in order to achieve high levels of excellence.

The study of the judging process in RG is not limited to the evaluation of the gymnast's final score but also the analysis of the RG judge's performance. Their capacity to give an accurate score to each gymnast routine during an entire competition is the first step to have a fair competition.

The high degree of attention required by the gymnast evaluation, applying the Cop requirements in such a short period of time can lead to some miss accuracy in judging the routines.

The aesthetics and art plays a major role in Rhythmic Gymnastics and consequently in its evaluation process. In the view of Baitsch (1974) and Iguatemy (1998), the characterization of the sport as an aesthetic phenomenon depends on the broadness of the concept of aesthetic. Besides many other factors there is the question of objectivity and subjectivity of aesthetics. This concept applied to the judging process in RG, making the evaluation of the RG routines a very complex task and can influence judge's performance in competition. This complexity is further emphasized by the

dynamic characteristic of the Cop that allow that gymnasts to create new elements (not listed in the CoP) which increase even more the difficulty of the evaluation process (Bobo & Sierra 1998). Whether the study of routines content or the way this content is understood and evaluated by coaches and judges are an essential task for the knowledge of RG sport (Laffranchi, 2005).

The main determinant of success in Rhythmic Gymnastics competition is the capacity to execute with maximum correction high level elements of body and apparatus technique, in perfect harmony with the character and rhythm of the music. From this starting point should result choreographies that by its originality and diversity present itself as a magnificent artistic spectacle for those who watch and be a challenge to high level performances. The load volume of the competition routines can be determined by the quantity, quality and diversity of the composition parameters (Arkaev, 2004). The result of the analysis of these factors may, according to Ávila-Carvalho (2012), influence the development programs in use, as well as de experimental designs used in scientific research in RG. The study of the content of competition routines presents a large amount of indicators with great representative and interpretative power by science as well as by practice.

The knowledge can also contribute to preview and characterize the effort requirements allowing improvements in the gymnasts' preparation to the competition readiness (Ferreirinha, 2009).

Ferreirinha (2009) refers that to determine the training models it is important to know the characteristics competition routines for high level gymnasts including the details concerning the specificity of their components.

In this context, we consider that the analysis of the evaluation variables that are reflected in the judgement and the analysis of the variables of technical content of the routines, both related, can contribute to a better understanding of the RG evolution process and also to develop the capacity to generate new developments in the sport. These complexity of the GR judging tasks, the difficulties experienced in acting as judge in RG high level competitions and the study of routines content or the way this content is understood and evaluated were our main motivation for these studies.

1.2. Main problem and goals of the thesis

Rhythmic gymnastics has been a sport in constant evolution due to a permanent analysis of the factors that affect it. The variables of judging in RG and the technical abilities (difficulty) of the routines of world reference gymnasts were set as the main points of our studies because we consider them as the key points to improve the quality of the high level competitions in the future.

Thus, before suggesting future changes, it is important to understand how it works in the present, finding out what should be changed and what should be kept.

Then the following questions were defined:

1) How does the performance of judges is effective to the competition success?

2) How does the structure of the evaluation tool (CoP) used by judges is effective in the optimization the competition success?

3) How does the quality of the judging is effective in the optimization the competition success?

4) How does the CoP requirements and the routine content seen/understood by the judges and coaches for effective in optimizing the success in competition for the high level RG individual gymnasts?

5) How the choose of the body technique elements included in the routines, is effective in optimizing the success in competition for the high level RG individual gymnasts?

Then we could define the main objective:

Analyse if the performance of RG judges and the quality of their judgement in competition, the accuracy of evaluation tool (CoP) and the content of high level RG individual gymnasts competition routines are effective and **contribute** to the success in competition.

Based on the main objective the following specific objectives were defined:

1) To analyse the accuracy of the judges' evaluation of the difficulty elements performed by the RG high level individual gymnasts in their competition routines.

2) To analyse the evaluation tool (CoP) from the perspective of the judges.

3) To analyse the quality of judging within the group of judges who evaluated a high-level RG individual competition.

4) To analyse how judges and coaches understand the CoP requirements and the routine content in the success competition.

5) To analyse the content of RG individual routines regarding the body elements and apparatus elements, according to the gymnasts' performance in competition.

1.3. Structure of the Thesis

This thesis is presented according to the rules and guidelines for dissertations presentation of the Faculty of Sciences of Sport and Physical Education of the Coruña University. The structure refers to the Scandinavian model composed of scientific articles published and submitted in journals with peer review.

The option to organize the thesis in different studies came from understanding that this model would allow the access in a progressive and sustained way to a deeper knowledge of the matters under review. The sequence adopted in the organization of the studies allowed gradually an increase of the autonomy and consistency of knowledge which opened new views for each subsequent study. At the same time, the preparation of studies, as well as their submission and their publication helped to widen the critical and debate space in the scientific community about the subject in focus.

Each study led us to the next study allowing broaden the range of knowledge that was gradually acquired, giving consistency and coherence to the work done

In consequence of the model choosen, the different parts of this work, related to the articles submitted and published in different journals, respect the rules of publication regarding the references rules of each journal. In this way we tried to approach the articles compiled here with those published by each journal.

Concerning the macro-structure, this thesis is organized into five chapters, which correspond to the stages of development and that allowed to respond adequately to the formulated questions and the to the proposed objectives.

Chapter I includes the Introduction that presents the framework of the studies, their justification and relevance, as well as the proposed objectives and the structure of the dissertation.

Chapter II integrates four empirical studies regarding of the evaluation in high level RG competition, which in gymnastics we call judging. In the first study we analyse the accuracy of evaluation done by judges regarding the difficulty elements performed in the RG high level individual routines. The second study analyses the evaluation tool structure (FIG, 2012) based on the perception of the judges. The third study analyses the quality of judgement within the group of judges who evaluate the gymnasts' performance. The last study this chapter is focused in the comparision of the coaches and judges view for the same routine content.

Chapter III includes two empirical studies. In these studies we analyse the content of the individual routines regarding the body difficulties, the performance levels in competition and the the apparatus used in the routine (Hoop, Ball, Clubs and Ribbon).

Thus we tried to map the empirical research focused on RG international judges performance and on the evaluation tool accuracy on one hand, and in another hand on the RG individual routines content trying to find an array which features the EXCELLENCE standards in the sport. Additionally we aimed to better understand the trends of research, without neglecting the consequences it brings to the competition field.

1.4. Summary of the studies presented on the Thesis

In the table 1 we present a summary of the studies done with the authores, the publication/submission date, the title and the journal where they were published or submitted.

Table 1. Summary of the studies presented on the Thesis.

Chapter II	**Empirical Studies - The judgment in Rhythmic Gymnastics high level competition routines**
Study 1	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce, M. (2015). Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines. Science of Gymnastics Journal. 7(3), 81-93
Study II	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce, M. (2016). The Evaluation Rules in the View of the Rhythmic Gymnastics Judges. Journal of Sports Science 4 (2016) 232-240
Study III	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce,M. (2016). Judging in Rhythmic Gymnastics at different levels of performance. Human kinetics Journal. (Submited)
Study IV	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce, M. (2016). Departure Difficulty Score Vs Final Difficulty Score. The effect of Performance in Elite Rhythmic Gymnastics. Athens Journal of Sports. Vol. X, No. Y
Chapter III	**Empirical Studies – Content of the Rhythmic Gymnastics high level competition routines**
Study V	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce, M. (2016). Technical Content of Elite Rhythmic Gymnastics. Science of Gymnastics Journal. 8 (1), 85-96.
Study VI	Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., Bobo-Arce, M. (2016). Quantitative and qualitative analyses of the Rhythmic Gymnastics Individual Routines in the different apparatus: Variety and Diversity. Revista Apunts Educatió Fisica e Esports. (Accepted for publication 20 april 2016)

Altogether, with these studies this dissertation aimed to gain understanding about the development of expertise among success and excellence on *Rhythmic Gymnastics,* in particular, about factors that may be related with the promotion or inhibition of expertise in the specific context of International highlevel sport. From this perspective, this dissertation ends in the Chapter IV and V, with a general discussion, as well as some practical implications of the present research to the field of judgement development, and to the field of success in the competition.

The final references for all of the studies are presented at the end of the whole thesis.

2. Empirical Studies: The judgment in Rhythmic Gymnastics high level competition routines

2. Empirical Studies: The judgment in Rhythmic Gymnastics high level competition routines

2.1. Study I: Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines

Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines.
Leandro, C. [1,3], Ávila-Carvalho, L. [2], Sierra-Palmeiro, E. [3], Bobo-Arce, M. [3]
Science of Gymnastics Journal 2015; 7(3), 81-93

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

2.1.1. Abstract

The main goal of this study was to analyse the accuracy in judging the Difficulty score in the Rhythmic Gymnastics Kiev World Championship 2013. The accuracy was determined analysing the judges' agreement on the evaluation of the routines difficulty elements. 1152 difficulty forms concerning 288 individual routines were analysed - 4 forms per routine, 1 per judge.

To allow the comparison between gymnasts with different levels the individual routines were clustered into 3 subgroups according to their final ranking competition. Body difficulty elements were organized, according to the composition requirements stated in the RG Code of Points (FIG, 2012). Non-parametric tests - Cochran's Q and Chi-Square Tests were applied to determine whether there were significant differences between groups. As main results we can point out that in general the judges did not agree on difficulty evaluation in 40% of the elements. The level of accuracy was lower in the second part of the ranking, and in the Mastery and DER difficulty elements.

Key Words: Evaluation, accuracy, judge, rhythmic gymnastics.

2.1.2. Introduction

Rhythmic Gymnastics (RG) is characterized by the high level of difficulty of the body elements and apparatus handling, combining esthetical and artistic components. This complexity increases the difficulty of the judgment and its accuracy mainly in high level performances. The requirements are quantitative (amount and variety of body and apparatus movements) and qualitative (degree of difficulty and quality in performance) and they are evaluated by the judges according to the rules and evaluation criteria stated in RG Code of points (Bobo, 2002).

The Body and apparatus movements are grouped according to the type of skills, the level of difficulty and the complexity of the movements (Lebre, 2011). The main groups considered in the routines evaluation are: Jumps, Balances and Rotations, Mixed difficulties, additional criteria for the body movements - waves and pre-acrobatics, Dance Steps, Mastery (special apparatus handling) and Dynamic Elements with Rotation and throw (DER).

In competition the performance is evaluated by 2 panels of judges: the difficulty (D) jury that judges the routines content and the execution (E) jury to evaluate the quality of the routines. The gymnasts present in each competition a difficulty form with all difficulties listed. Each judge must confirm the difficulty elements performed by the gymnast and cross out those that are not correctly performed or not performed at all (FIG, 2012). The final D score is the average of two intermediate scores. When the score become published on the screens, the judges can compare the final score to their own scores. Therefore, the judges score independently although there's still some feedback (Bucar, Cuk, Pajek, Kovac, & Leskosek, 2013).

In previous studies was noted that judging is not only a matter of identifying the sports performance. There are also various facts, identified in the literature, having an influence on the several stages of processing information in gymnastics judgment (Leandro, 2009). Findlay and Ste-Marie (2004) found out that the were the judges tend to judge better the gymnast higher qualified in previous competitions, concluding that the reputation of the gymnasts have influence on the judging. The judge's experience has been also described as influencing the quality of judgment. Leandro, Ávila-Carvalho, and Lebre (2010) and Ste-Marie, Valiquette, and Taylor (2001) found that the more experienced judges had better perception and anticipation of the elements and there

for, were better evaluators. Other factors, as the memorizing capacity (Ste-Marie, Valiquette, and Taylor, 2001), and the tendency to adapt their scores to those given by the judges of the same panel (Boen, Karen, Yves, Jos, and Tim, 2008) were also described. The observation angle (Plessner and Schallies, 2005) and the judges with experience as gymnasts (Heinen, Vinken & Velentzas, 2012) were also described as factors that can influence in the judges accuracy. Besides these factors, is also relevant to know whether the factors related to the sport specificity as the structure/organization of the Code of Points, the evaluation criteria defined by the sports authorities has an influence (positive or/and negative) on the judge's performance and consequently on the gymnasts final scores. Rhythmic gymnastics has been experiencing a constant and outstanding evolution in its' technic for the last few years because of the evolution of the Code of Points (Palomero, 1996). The evaluation of the gymnasts is made by a collective observation of judges that should be objective. However, this evaluation is not yet exact, probably due to huge amount of evaluation criteria defined for each difficulty element. This can be verified by the differences registered between the judges of the same panel when the evaluate the same routine. This fact is wellknown in the sport but not yet studied. The majority of studies available deal with the analysis of the technical content of exercises or with the final scores given at the end of each exercise. We could not find any study dealing with the analysis of the difficulty evaluation, element per element, trying to see if the final score of each judge are the product of the validation of the same difficulty elements. Under this subject, the most relevant studies we found are Palomero (1996) and Bobo (2002), in which both the authors present a new proposal for the scoring, based in performance indicators. Čuk, Fink, & Leskošek (2012) studied the way the different type of final score calculation can change the gymnasts final ranking. Gambarelli, Laquinta & Piazza (2012) developed a formula to avoid pre-agreements between judges. They proposed that the score from the judge of the same country of the gymnast should not enter in the calculation for the gymnast final score. Furthermore, they consider that this would be a factor of guarantee of higher reliability of the final score.

Some of the studies demonstrate that the structure of the Code of Points itself holds decisive influence in scoring gymnasts. In this way is very important to suggest alternative evaluation tools that respect the principles of evaluation (objectivity,

validity, reliability, discriminating power and practical utility) and allow a balanced appreciation of the different dimensions of the sport, in either aspects of quality or quantity in the performance of gymnasts (Bobo, 2002).

On the other side, the permanent changes in the Code of Points may cause a lack of understanding of the rules, which lead to a need of evaluation of judging instrument itself (Kirkpatrick & Hawk, 2006). Mark & Shotland (1987) remarked, any evaluation model has to be based on a group of principles, axioms and postulates that must be feasible. To have a Code of Points with an extremely complex model of evaluation that does not work when it has to be used, must be avoid. According to Bartolomeis (1999) it is not possible to see everything at the same time. The essential point is that the evaluation instrument evaluates what it is supposed to evaluate. For Tamir (1998) the evaluation criteria used should be tested in both validity (precision) and reliability (internal consistency). We could not find any study based on the analysis of the judges' activity based on the using of the difficulty forms during the competition, making this study a pioneer in this field.Thus, before suggesting future changes, it is important to understand how it works in the present, finding out what should be changed and what should be kept. According the pyramidal structure of the evaluation process (Figure 1) we established the goal of the study.

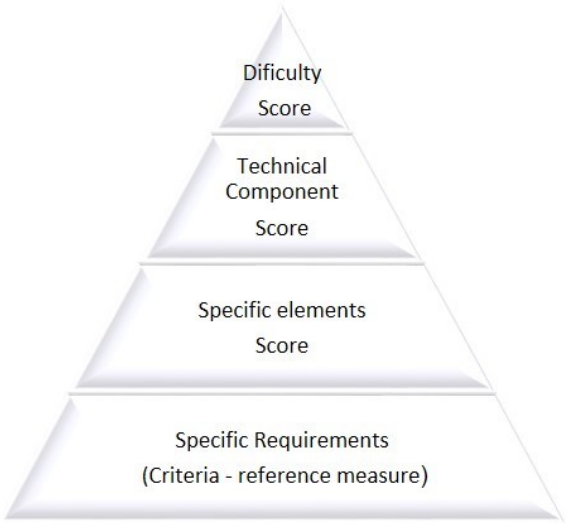


Figure 1. Pyramidal structure for analysis of the evaluation process.

The goal of this study was to analyse the accuracy in judging Difficulty in the Kiev World Championship 2013, trying to learn if the 4 difficulty judges evaluate in the same way the difficulty elements on the D form (agreement between the 4 judges). This accuracy was studied for each element declared in the difficulty form trying to understand if the perception of the validation criteria for each elements is similar for all judges. The final difficulty score given by each judge to the same gymnast were very similar, but, with this study, we will analyse if the judges arrived to the final score validating the same elements or validating different elements. After analysing the date in a global way, we will study the level of agreement between the judges concerning the validation of the difficulty elements according to: (1) the position of the gymnast on the final ranking (1st part, 2nd part and 3rd part), (2) the routine apparatus (hoop, ball, clubs and ribbon), and (3) the type of difficulty element.

2.1.3. Methods

Subjects and design

1152 difficulty forms concerning 288 individual routines were analysed (4 forms per routine, 1 per judge). The routines were performed by gymnasts from 45 different countries competing at Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013. This study was done with the permission of the International Gymnastics Federation. Full blinding of the judges involved was undertaken.

All difficulty elements reported in the difficulty forms provided by the gymnasts at the competition were analysed. Each element was considered validate or not according the notes done by the judge on the form. For each element, we studied the cases of agreement when all 4 judges validate or not the difficulty element and the disagreement when at least one of the judge did not validate and the others consider the element correctly done. The analyse was done considering the all sample, and the sample clustered into 3 subgroups according to gymnasts final ranking as follows: the first part of the ranking - the top 24 gymnasts, the second part of the ranking - 24 middle gymnasts and third part of the ranking – the 24 lower placed gymnasts on the ranking, to allow the comparison the agreement level of the judges when they evaluate gymnasts with different levels. Then, we studied the sample according to the apparatus used to perform the routine (hoop, ball, clubs and ribbon), and the type of difficulty element

performed listed according to the composition requirements of the Code of Points (FIG, 2012), (Figure 2).

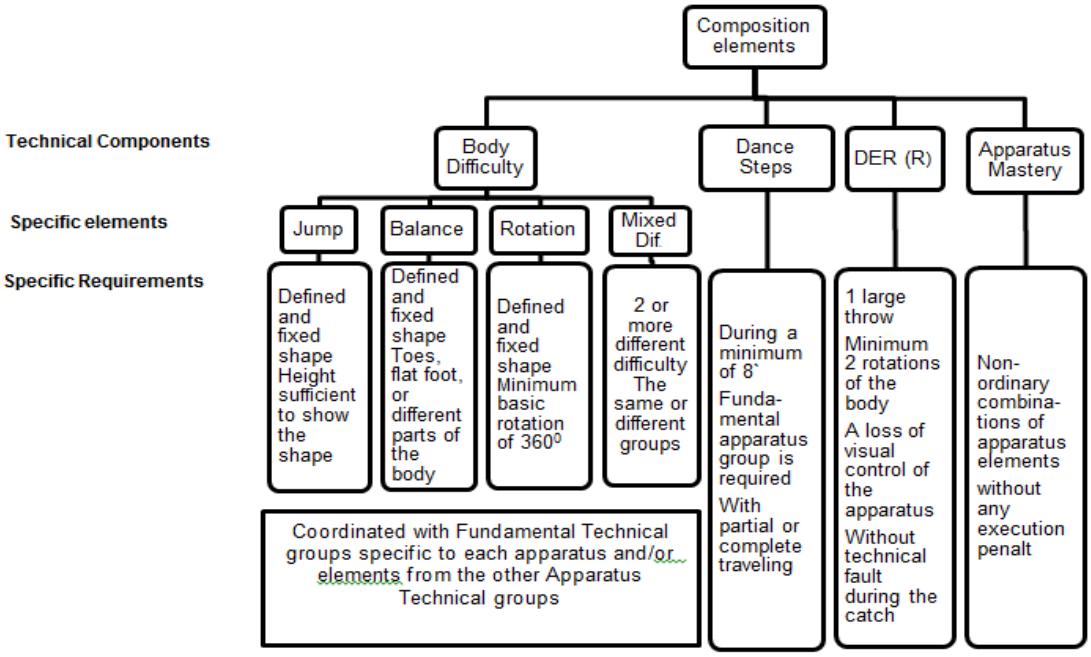


Figure 2. Technical Content of Rhythmic Gymnastics of Individual Gymnasts Routines (COP 2012/2016).

Statistical Analysis

For the statistical analysis we used the Statistical Package for the Social Sciences - Version 21.0 (SPSS 21.0, Chicago, USA) and Microsoft Office Excel 2010. Non-parametric tests (Cochran's Q and Chi-Square Tests) were applied to determine if there were significant differences between groups. We use the Chi-square Tests for two independent samples to study the differences between two groups for each variable and the Cochran`s Q test to analyse when a set of K differs significantly. Significance level was set at $\alpha = 0.05$ (corresponding to a confidence level of 95%).

2.1.4. Results

The forms were analysed first in a global way. For each difficulty element presented on the forms, the percentage of agreement between the 4 difficulty judges concerning the evaluation of the elements was determined. Then, the level of agreement on the elements evaluation was also calculated with the sample dived in 3 groups according to the final ranking of the gymnasts (Table 1).

Table 1. Level of agreement on the evaluation of the difficulty elements presented on the D Forms for all sample, and for the 3 groups according to the final ranking of the gymnasts.

	All Sample		1st part of the Ranking		2nd part of the Ranking		3rd part of the ranking	
	n	%	n	%	n	%	n	%
Not Agree	4871	40.0	1300	31.2	1836	43.9	1735	45.4
Agree	7294	60.0	2865	68.8	2343	56.1	2086	54.6

Chi-Square Test (Asymp.Sig.(2sided)) .000 * (*P<0.05)

The judges agreed on the evaluation of 60.0% of the difficulty elements presented on the difficulty forms. When we observe the results according to ranking of the gymnasts, is visible that higher the gymnast is placed in the ranking, higher is the agreement of the judges on the difficulty elements evaluation: 68.8% on the first part of the ranking, 56.1% on the 2nd part and 54.6% on the 3rd part. According to the results of the Chi-Square test, the differences between the cases of agreement and disagreement on the evaluation of the difficulty elements were statistically significant in all cases. Studying the difficulty forms according to the routine apparatus (Table 2) we observed that the range between the disagreement values for the elements evaluation in the 4 apparatus is not very wide (from 37.3% in ball to 41.2% in hoop). However, when we observed the results of the Chi-Square test we could verify that for all apparatus there were significant differences between the values of the agreement and the disagreement on the evaluation of the difficulty elements.

Table 2. Level of agreement on the evaluation of the difficulty elements presented on the D Forms according to the routine apparatus.

	Hoop		Ball		Ribbon		Clubs	
	n	%	n	%	n	%	n	%
Not Agree	1370	41.2	1129	37.3	1191	41.0	1244	40.6
Agree	1867	58.8	1894	62.7	1715	59.0	1818	59.4

Chi-Square Test (Asymp.Sig. (2sided)) .000 * (*P<0.05)

Comparing the data between apparatus through the Cochran`s Q test (Table 3) we could find that there is a significant difference between the values registered for Hoop and Clubs (p value 0.018 and 0.000 respectively), what showed that there was differences in judges agreement level on the elements evaluation for the different apparatus.

Table 3. Results of the Cochran`s Q test comparing the results the agreement level for Hoop, Ball, Clubs and Ribbon routines.

	Hoop	Ball	Clubs	Ribbon
N	3174	3023	3062	2906
Cochran's Q	9.960	6.512	25.174	6.232
Sig.	.018*	.090	.000*	.099

(*P<0.05)

Continuing the analysis in each apparatus, we studied the lack of agreement between judges regarding the final ranking of the gymnasts.

Table 4. Results of the Cochran`s Q test comparing the results the agreement level for Hoop, Ball, Clubs and Ribbon routines according to the final ranking of the gymnasts.

	Hoop			Ball			Clubs			Ribbon		
	1st part	2nd part	3rd part	1st part	2nd part	3rd part	1st part	2nd part	3rd part	1st part	2nd part	3rd part
N	1069	1078	1027	1044	1036	943	1050	1061	951	1002	1004	900
Cochran's Q	5.167	22.273	2.385	10.793	6.660	6.281	7.482	16.485	4.821	18.351	10.042	5.405
Sig.	.173	.000*	.499	.013*	.083	.095	.061	.001*	.185	.000*	.019*	.145

(*P<0.05)

The results of the Cochran`s Q test (Table 4) revealed that in Hoop, and Clubs the judges disagreed significantly only on evaluation the difficulty elements of the gymnasts ranked in the 2nd part of the final ranking; in Ball they disagree significantly on the gymnasts in the 1st part of the final ranking; and finally for Ribbon they disagree significantly on the 1st and 2nd part of the final ranking.

We studied the level of judges agreement on the difficulty elements considering the different group of elements described in the Code of Points (Table 5).

Table 5. Level of agreement on the evaluation of the difficulty elements presented on the D Forms according to the different type of elements.

	Not Agree		Agree	
	N	%	N	%
Mastery	726	62.5	436	37.5
Dance Steps	220	28.7	546	71.3
DER	1871	40.6	2735	59.4
Jumps	270	35.6	489	64.4
Balance	302	43.1	398	56.9
Rotations	1065	32.0	2263	68.0
Mixed Difficulties	93	38.3	150	61.7
Criteria assoc. to diff.	324	53.9	277	46.1

Chi-Square Test (Asymp.Sig.(2sided)) .000 * (*P<0.05)

In the most part of the groups of elements the agreement percentage between the judges was higher than the disagreement percentage. Only for the evaluation of the Mastery group and the criteria associated to the difficulties (waves and acrobatic skills) the percentage of disagreement between the judges was higher than the agreement - 62.5% and 53.9% respectively for the agreement against 37.5% and 46.1% for the disagreement. Despite this remark, the results of the Chi-Square test the differences between the cases of agreement and disagreement on the evaluation of the difficulty elements were statistically significant in all cases.

The level of agreement between the judges evaluating the different groups of elements was, then, studied regarding the final ranking of the gymnasts (Table 6).

Table 6. Results of the Cochran`s Q test comparing the results the agreement level for different groups of elements according to the final ranking of the gymnasts.

	1st part			2nd part			3rd part		
	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.
Jumps	257	1.227	.817	244	4.483	.208	258	2.92	.401
Balances	207	1.224	.785	238	5.89	.121	255	6.084	.106
Mastery	361	116.05	.000*	394	46.744	.000*	407	32.992	.000*
DER	1607	62.548	.000*	1567	8.492	.036*	1432	17.251	.001*
Dance Steps	260	8.12	.047*	244	14.709	.002*	262	2.121	.551
Rotations	1168	56.937	.000*	1185	1.625	.652	975	4.288	.224
Mix. Diff.	108	10.553	.015*	81	10.881	.012*	54	8.937	.030*
Criteria	197	12.425	.005*	226	5.158	.164	178	3.774	.282

(*P<0.05)

Observing the results we can see that for Jumps and Balances was not remarked a significant disagreement between judges on the evaluation of the elements performed by the gymnasts independently of their placement in the final ranking. For the Dance Steps, there was only a significant disagreement between the judges for the gymnasts placed in the first and second parts of the ranking. Regarding the Rotations and the Criteria associated to the difficulties the significant disagreement was registered only for the gymnasts placed on the first part of the ranking. When we observe the Table 6, we can see that there are statistically significant differences for the Mastery elements, the DER elements and Mixed Difficulties in the 3 parts of the ranking, once the p value are null or extremely low, what shows clearly the disagreement between the judges.

For the analysis of the rotations we divided them in 3 sub-groups (RPIV - *relevé* rotations (pivot), RFF - rotations on the flat foot or on other part of the body and RF - *fouetté* rotations), because of their different characteristics that means different evaluation requirements (COP, 2012). In each sub-group of RPIV and RFF rotations, we analysed separately the basis of the rotation and the number of rotations associated to the basis.

The level of agreement of the judges evaluating the different type of rotations elements was, then, studied regarding the final ranking of the gymnasts (Table 7).

Table 7. Results of the Cochran`s Q test comparing the results the agreement level for different groups of rotations elements according to the final ranking of the gymnasts.

	1st part			2nd part			3rd part		
	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.
RPIV Base	195	2,769	.586	167	2,780	.448	188	7,554	.051
RPIV Rotations	431	37,748	.000*	346	1,213	.763	333	0,283	.969
RFF Base	99	2,314	.594	96	4,116	.295	65	7,627	.050*
RFF Rotations	273	11.634	.008*	198	3,915	.283	106	7,382	.060
RF	206	13,481	.004*	378	7,774	.050*	283	2,928	.419

On the Table 7 we can see that for the basis of RPIV and RFF, there is no statistically significant difference between the evaluation done by the judges in the first and second parts of the ranking. We can see, also, that the values for significance drop substantially in the third part of the ranking. When we analyse the rotations associated to the basis of RPIV and RFF, we can see that in the first part of the ranking that the p

value shows clearly the disagreement between the judges in evaluating such part of the difficulty.

Concerning the *fouetté* rotations, there is no agreement between the judges in the first and second parts of the ranking.

2.1.5. Discussion

The goal of this study was determine the accuracy of the judges on the evaluation of each difficulty element presented in the difficulty forms.

Studying the forms in a global way we found that the percentage of elements where the 4 judges of panel agreed on the elements evaluation was higher than the disagreement cases. Nevertheless, we could observe that the judges agreed only in 60% of the elements, what is not enough for an evaluation that is supposed to be exact and accurate.

When we divided the gymnasts in 3 groups according to their place in the final ranking we found out that the judges showed a higher percentage of agreement for the gymnasts placed in the first part of the ranking and lower when we went down through the ranking. These results may suggest that it is more difficult for the judges to evaluate with precision the average and low level gymnasts. This evidence might be related to some criteria to validate the elements that, probably are not enough specific, what can cause some pliability in the evaluation. To solve this problem Simões (2000) suggests that all evaluation systems should hold precise criteria to allow judging correctly the performance. When the gymnast performs perfectly or almost perfectly the element, as usually happens with the top ranked gymnasts, is easier to the judges to recognize the difficulty, applying the evaluation criteria clearly, and tend to agree on its the evaluation. According to Bartolomeis (1999), the evaluation criteria are defined based on a successful criteria, which can facilitate the agreement of judges when the gymnast perform the elements with success, which is the case for the top ranked gymnasts. For the average and low level gymnasts is clearly more difficult to determine the “drop off” point to validate the difficulty elements because these gymnasts are doing the elements with some technical faults which leads the judges to struggle in applying the evaluation criteria stated in Code of Points (FIG, 2012).

We can also speculate that there could be an influence from what is expected, once the judges might expect better gymnasts to perform the difficulty elements correctly, as Findlay & Ste-Marie (2004) found, in a study with figure skating performances, that the judges gave higher scores to the better known skaters, comparing to the less known ones.

Other point that should be added to this discussion is the fact that the evaluation criteria for some difficulty elements include, according to the Code of Points (FIG, 2012) some points concerning the quality of execution that may contribute to a higher variability on the validation of the elements. The interference of these execution quality criteria may create some variability in the work of the difficulty judge, creating some "grey zones" in the evaluation of difficulty elements. According to Askew (2002), the evaluator should direct all his attention for a specific profile and ignore the interference of any other information from a different profile.

The analysis of the results by apparatus revealed that the percentage agreement had not big differences for the routines performed with different apparatus. The results showed that behavior does not change from one apparatus to another; on the contrary we could remark that there was a consistency on the lake of accuracy in the difficulty elements evaluation. This consistency is due to the fact that the difficulty elements used in the different apparatus are basically the same and therefore, the requirements to validate the apparatus are the same (FIG, 2012).

Observing the results obtained for the judgment accuracy when we studied it for each apparatus and according to final ranking of the gymnasts we found out that the lower values of accuracy in the judgment were registered mainly in the gymnasts of the second part of the ranking. Besides what was already discussed about the lack of precision in defining the evaluation requirements, we are still able to speculate about the short amount of time that each judge has to consider a great amount of requirements defined for every single element in the routine composition, which may cause high variability between judges scores (Čuk & Karacsony, 2004). This is a problem for the average gymnasts because in opposite to higher level gymnasts where is easy to identify the difficulty elements correctly done and to lower level gymnast where visible when they do not perform the difficulty elements correctly, the average gymnasts often

present an unclear version of the difficulty element making the decision to validate an element even more difficult than usual.

The results obtained when we analysed the level of agreement of the judges according to the type of difficulty element evaluated showed that the judges could not agree on the evaluation of the Mastery elements, and the Criteria (waves and pre-acrobatic elements) associated to the difficulty elements. These two groups showed levels of disagreement higher than the agreements, clearly in opposition to what happened with the other groups. The results suggests that definition of the evaluation requirements may have not an enough clear statement in the Code of Points (FIG, 2012), which can lead the judges in troubles to decide when the elements should be validate or not. According to the technical requirements to validate a Mastery element, it should be "a combination of extraordinary apparatus elements performed without technical faults". The definition of "extraordinary apparatus elements" is too vague to allow the judge to evaluate the elements with accuracy and could be also influenced by the international experience of the judge: after judging a certain number of international competitions the level of expectation for an "extraordinary element" can be raised. Knowing that in the World Championships the judges (one for each country participating) has different background experiences, we can understand that they cannot evaluate this technical requirement with same level of accuracy. In this way we strongly recommend that the Code of Points should include much more precise definitions of the technical requirements, because, according to Simões (2000) the evaluation criteria should be understood in equal manner by the various evaluators, in a way that the effect of the evaluation done may be valid and reliable.

After a more detailed analysis of each group of difficulty elements according to the gymnasts ranking, we could see that for the Jumps and Balances the level of agreement between judges was similar in the 3 parts of the ranking, showing that in these elements the judges apply the same evaluation criteria. The evaluation criteria are understood and applied in the same way by the evaluators, once they produce the same result. This result allows us to speculate that visual image of the element allows a quicker and more reliable understanding, once the stated difficulties are presented. Boen, Karen, Yves, Jos, and Tim (2008) reach the conclusion that the possibility of feedback creates agreement between gymnastic judges. We know (unpublished study), that

jumps and balances are repeated frequently in exercises, by the gymnasts in different apparatus routines, what facilitates the visual experience of the judge and therefore more precision in the application of evaluation criteria. According to Ste-Marie, Valiquette, & Taylor (2001), the visual image that is kept in the memory can influence the judge's performance. The agreement may be higher in the elements that appear often in exercises and because of that the judges have a clearer visual image and therefore a more precise evaluation.

In opposition, we can see that there are statistically significant differences between the 3 parts of the ranking in Mastery elements and DER elements, what clearly reveals the disagreement between the 4 judges on the validation of these difficulties. Besides what was already discussed above about the validation of Mastery elements, it is still relevant underline these elements are not listed and therefore the higher number of possible combination of handling contribute to make the evaluation of these type of elements even more difficult. We understand here that the absence of a list of Mastery elements would bring high improvements in routines creativity, although this could also bring the possibility for mixing originality concepts that should and must be evaluated in the originality item stated in COP (FIG, 2012). According to Balcells, Martín & Anguera (2009) it is possible to evaluate the originality and creativity with validity and reliability defining evaluation criteria that can be seen by the evaluators.

In the case of DER elements, the results lead us to the high number of criteria to bear in mind for the judge during the observation. According the Code of Points (FIG, 2012), the DER has an unlimited value and may contain till 19 different criteria that can be repeated. The judge has to memorize the criteria done to have the possibility to cross out on the difficulty form those what were not performed correctly or not done at all. Ste-Marie and Lee (1991) and Ste-Marie, Valiquette, & Taylor (2001) showed that the objectivity of a judge can be compromised by biases of memory. Also, the high number of criteria performed in such short may be responsible for this lack of agreement between the judges. We can speculate that the small amount of time that the judge has to observe and make all the possible deductions on the Difficulty form could be other source of variability between judges which may cause the evaluation of this group more vulnerable. Bucar, Čuk, Pajek, Kovac, & Leskosek (2013) and Čuk & Karacsony (2004)

identified this same problem in the evaluation of the Vault execution in female artistic gymnastics, once this is also done in few seconds with 21 possible deductions.

The data concerning the Dance Steps showed also a significant disagreement of the judges in the validation. Dance Steps has, as criteria to be validate, the duration of at least 8 seconds, which can cause high variability in the evaluation, since this evaluation is done without a stopwatch or other device, but through the sensibility of the judge, and can be serious influenced by the *tempo* of the music.

The evaluation of the Mixed difficulties and Criteria associated to the difficulty elements (acrobatic elements and waves) reveals a significant disagreement between the judges, which could be due to the statement on the Code of Points concerning the link between the wave or acrobatic element and difficulty element itself. According to COP (FIG, 2012) the link must be immediately before or after but it is not clearly specified if it should be in continuity of the difficulty element or if it could be a composition of two elements. According to Plessner (2005), the non-stated rules which can be considered as social norms, may influence the judge's decisions. It's important that they have great knowledge of the rules, to avoid wrong decisions.

Concerning the rotations, we can see that when evaluated the base of RPIV and RFF, there's no significant difference in the evaluation, in the first and second parts of the ranking. However, we can see that the values of a significant decrease in the third part of the ranking. Normally it is on the third part of the ranking where we find the lower level gymnasts and therefore with poor execution technique straight from the base of the rotation. According to the COP, the judge has to see the form, the degrees (360º) of the first turn and the technical faults that cancel the difficulty. The junction of all this factors (which are more present in the lower level gymnasts) belonging to two different profiles (difficulty and execution), may be explain the results of variability between judges found in the evaluation of this part of the difficulty.

Concerning the number of rotations associated to the base of RPIV and RFF, we can see that in the first part of the ranking there is clearly disagreement between the judges in evaluating these difficulty elements. About *fouetté* rotations, we found that in the first and second parts of the ranking there is no agreement between the four judges.

It is in the first and second parts of the ranking that the rotations performed done have a higher number of turns. By the evaluation criteria stated in COP, the judge has to

count the number of full turns performed that is sustained fixed, without technical faults. Then, the difficulty in counting a high number of turns performed (that can go upper than 10 turns, mainly in *fouettés*) at high speed in few seconds, identifying the technical faults that implies the cancellation of the difficulty, may be in the origin of this variability for this kind of elements, in the first part of the ranking. Once again, we highlight here the interference of some criteria concerning execution, when judges are judging difficulty. According to Plessner (2005), positive and negative effects of prior knowledge on referee decisions and observation of a high amount of demand in such a short amount of time, may cause the loss of important information.

2.1.6. Conclusion

The four judges of difficulty panel did not agree in their evaluation in 40% of the difficulty elements presented in the difficulty forms. Regarding the final ranking of the gymnasts the agreement level is higher in the high and low level gymnasts. The level of accuracy was lower in the second part of the ranking, and in the difficulty elements which validation criteria depends not only from difficulty criteria but also from execution criteria.

The analysis by type of difficulty elements showed that for the Jumps and Balances the judges agreed on the evaluation of the elements which means an acceptable accuracy of judgement, but for the other types of elements the level of disagreement between the judges was significantly high to be an accurate judgement, where we highlight the Mastery and DER difficulty elements. This study provides updated information about the precision of difficulty judging in rhythmic gymnastics, to be considered in the possible alteration of the present code of points, in particular in the definition of the evaluation criteria of the elements where we see the highest disagreement between judges.

2.1.7. References

Askew, S. (2002). Feedback for learning. Journal of eduction for teaching, 28, 83-90.

Balcells, M., Martín, C., & Anguera, M. (2009). Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión Corporal y Danza Contact-Improvisatio. Apunts educación física y deportes.Ciencias aplicadas a la actividad física y el deporte, 14-23.

Bartolomeis, F. (1999). Avaliação e Orientação: Objectivos, instrumentos e métodos. Lisboa: Livros Horizonte.

Bobo, M. (2002). El juicio deportivo en Gimnasia Rítmica. Una propuesta de evaluación basada en indicadores de rendimiento. (PHD Thesis), Universidad da Coruña, Instituto Nacional de Educación Física de Galicia, Coruña.

Boen, F. B., Karen, H., Yves, V. A., Jos, F., & Tim, S. (2008). Open feedback in gymnastic judging causes conformity bias based on informational influencing. Journal of Sports Sciences, 26(6), 621–628.

Bucar, P. M., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? Journal of Human Kinetics, 37, 173-181.

Čuk, I., Fink, H., & Leskošek, B. (2012). Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. Science of Gymnastics Journal, 4, 73 – 82.

Čuk, I., & Karacsony. (2004). Vault: methods, ideas, curiosities, history. Ljubljana: STD. Sangvinck.

FIG. (2012). Code of Points for Rhythmic Gymnastics Competitions. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

FIG. (2013). Gymnastics Results. Available at: http/www.gymnasticsresults.com

Findlay, C., & Ste-Marie, M. (2004). A Reputation Bias in Figure Skating Judging. Journal of Sport & Exercise Psychology, 26 (1), 154-166.

Gambarelli, G., Laquinta, G., & Piazza, M. (2012). Anti-collusion indices and averages for the evaluation of performances and judges. Journal of Sports Sciences, 30(4), 411-417.

Heinen, T., Vinken, P., & Velentzas, K. (2012). Judging Performance In Gymnastics: A Matter Of Motor Or Visual Experience? . Science of Gymnastics Journal, 4, 63 – 72.

Kirkpatrick, J., & Hawk, L. (2006). Curricula and evaluation: Maximizing results. Measuring and Evaluating. Available from EBSCO

Leandro, C. (2009). Avaliação de Juízes de Ginástica Rítmica. (Master Thesis), Porto University, Porto.

Leandro, C., Ávila-Carvalho, L., & Lebre, E. (2010). The avaluation of the performance of Rhythmic Gymnastics` Judges. Palestrica of the Third Millennium Civilization & Sport, 11(3), 202-206.

Lebre, E. (2011). Technical principles for the new framework. Crossroads to the Future [Press release]

Mark, M., & Shotland, R. (1987). Multi Methods in Program Evaluation. New Directions For Program Evaluation (Vol. 35). Londres: Jossey – Bass Inc.

Palomero, M. L. (1996). Hacia una objetivación del Código Internacional de Gimnasia Ritmica Deportiva.Barcelona.

Plessner, H. (2005). Positive and negative effects of prior knowledge on referee decisions in sports. T. Betsch, Haberstroh,S (Ed.) The routines of decision making (pp. 311–324).

Plessner, H., & Schallies, E. (2005). Judging the Cross on Rings: A Matter of Achieving Shape Constancy. Applied Cognitive Psychology, 19, 1145-1145.

Simões, G. (2000). A avaliação do desempenho Docente. Lisboa: Texto Editora.

Ste-Marie, D., & Lee, T. D. (1991). Prior processing effect on gymnastic judging. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 126-136.

Ste-Marie, D. M., Valiquette, S. M., & Taylor, G. (2001). Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. Research Quarterly for Exercise and Sport, 72(4), 420-426.

Tamir, P. (1998). Assessment and Evaluation in Science Education . Opportunities to Learn and Outcomes International Hand book of science Education (pp. 761-789): Dordrecht: Kluwer Academic Publishers.

2.2. Study II: The Evaluation Rules in the View of the Rhythmic Gymnastics Judges

The Evaluation Rules in the View of the Rhythmic Gymnastics Judges.
Leandro, C.[1,3], Ávila-Carvalho, L.[2], Sierra-Palmeiro, E.[3], Bobo-Arce, M.[3]
Journal of Sports Science 2016; (4) 232-240

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

2.2.1. Abstract

The Code of Points (CoP) is the evaluation tool of Rhythmic Gymnastics (RG) and contributes to its evolution as sport. This tool is applied in competition by the judges who have a crucial role in the performance evaluation. The aim of this study was to characterize the International judges, their perception of the objectivity of their judgment in competitions and their opinions about the content and the reliability of the Code of Points. 162 international RG judges answered a questionnaire specially designed for this study. For the data analysis, non-parametric tests were used. According to the judges, the evaluation of the difficulty component in the individual routines has more subjectivity in the items of Mastery (58.6%) and Dance Steps (55.6%) and is less subjective in the body difficulties of Balance (72.2%). Concerning the execution, the judges consider that the evaluation of the artistic faults is the most subjective in this domain (64.8%). Within the artistic faults the items unity of the composition and relation between the Music and the Movements, were those which registered higher significant results for the subjectivity in the evaluation (47.5% and 37.0%, respectively).

Key words: Evaluation, objectivity, judge, rhythmic gymnastics.

2.2.2. Introduction

Performance the evaluation in RG depends on a judging process done by specialized judges that apply a specific set of rules and procedures established in the official FIG Code of Poins (CoP) [1]. The CoP is used as an evaluation instrument by the judges, who have their own technical, human and social experiences. All these aspects take part in the judging process and from it depends, in part, the improvement of quality of the sports practice, the safeness of the physical and moral integrity of the athletes and the reenforcement of ethical values in sport.

Thus, the judge's background is a factor of sportive quality. The performance of international level judges overcomes the performance of less experienced judges, once they use of other cognitive strategies, increasing their global efficiency of error spotting [2]. Also, "If we seek sports excellence, it is not possible not to understand the job of judgement and refereeing in sport. It is not enough to train coaches with scholar degrees, masters and PhDs, it's necessary to think with vision, because it is definitely incongruent and even not logic for the one who evaluates the process and final result of the coach work, not to hold an integral formation that amounts to the level in which the process is occurring" [3]. Academical, professional and social formation of the referee is a characteristic acknowledged by them as being a guarantee of their performances quality [4].

On the other side, the CoP which rules and orientates all the judges' actions, works as an evaluation tool that depending on its structure, its content, and its reliability. All this factors may have a better or worse impact on the judges' performance as evaluators. It is concluded that the most valued skills are those related to the sport's technical parameters and the ability to adapt to any level of competition with self-confidence and self-assuredness [5]. Considering that the judge and the CoP hold an inseparable dialectic, it is necessary to analyze them together. So the aims of this study are: to define the population of the "International RG Judges" according to personal information, education, professional experience and experience as judges; and to identify their opinion about the CoP 2012 related to its content, structure, clarity and validity of the rules to be applied, as well as possible changes to be introduced to contribute to and improvement of the judging and therefore the correct evolution of the sport.

2.2.3. Methods

Subjects and design

162 international RG judges answered a specific questionnaire specially developed for this study. It was composed by 15 questions grouped in 2 categories: (1) personal information, education, professional experience and experience as a judge, and (2) objectivity of evaluation in RG and proposals to change the Code of Points [1].

This study was approved by the International Gymnastics Federation (FIG)**.** All the international RG judges, from all over the worldw, 287 judges were invited by FIG to answer the questionnaire available at Google Drive.

To protect the judges' anonymity, the answers were received anonymously on google drive, so the full blinding of the judges involved was undertaken. The data was collected between July and September 2014.

Statistical Analysis

For the statistical analysis, we used the Statistical Package for the Social Sciences-Version 21.0 (SPSS 21.0, Chicago, USA) and Microsoft Office Excel 2010.

For the data analysis, non-parametric tests were used (Friedman test and Sign test between groups) to determine whether there were significant differences between groups. Significance level was set at $\alpha = 0.05$ (corresponding to a confidence level of 95%). The frequencies and percentages of the prognostic variables were calculated through the descriptive statistics.

2.2.4. Results

The characterization of the Judges is resumed in Table 1.

Table 1. Descriptive statistic of the judges characterization.

Characterization of Judges (*N* = 162)					
				Freq.	%
Personal information	Sex	Female		160	98.8
		Male		2	1.2
	Age	Mean	43.4		
		Minimum	22		
		Maximum	68		
	Country			59	
Education	High School			16	9.9
	University			80	49.3
	Master			51	31.4
	PHD			15	9.4
Work Experience (RG Coach)	Yes			141	87
	No			21	13
RG International Judge	Brevet I			6	3.7
	Brevet II			29	17.9
	Brevet III			66	40.7
	Brevet IV			61	37.7
RG International Judge Experience	Less than 1 Olympic cycles			42	25.9
	1-2 Olympic cycles			51	31.5
	More than 2 Olympic cycles			69	42.6

The judges were 43.4 years old, 49.3% have a university degree, 87% are also coaches, 40.7% are judges brevet III and 42.6% have been judges for more than 2 Olympic cycles.

We can see in Table 2 the summary of the collected data about the judges' opinion about the objectivity of Difficulty evaluation. The *Mastery* is considered to be the item with less objectivity in evaluation with the answer "less objective to evaluate" collecting 58.6% of the answers. Following we have, in increasing order for the objectivity in judgement, the following groups: *Dance Steps*, *DER (base)* and *Rotation*s in which the answer "more or less objective to evaluate" is the more frequent one 55.6%, 45.7% and 51.2%, respectively. In the groups *DER (criteria)* and *Jumps* the answer "objective to evaluate" is the more frequent with 40.7% and 46.3% respectively. Yet, it's in *Rotations (Basis)* and *Balances* that we see higher values of objectivity in the evaluation with the answer "objective to evaluate" getting 59.9% and 72.2% of the answers, respectively.

Table 2. Descriptive statistics of the judges' opinion about the objectivity in difficulty evaluation.

Objectivity of Evaluation	Mastery	Dance Step	DER (Base)	DER (Criteria)	Jumps	Balance	Rotation (Base)	Rotation (Added)
	Frequency Tables (%)							
Less objective	58.6	22.8	21.0	21.6	11.1	3.1	7.4	16.0
More or less objective	34.0	55.6	45.7	37.7	42.6	24.7	32.7	51.2
Objective	7.4	21.6	33.3	40.7	46.3	72.2	59.9	32.7
Teste de Friedman	$P = 0.000^{*}$							

Figure 1 presented the average indicator of objectivity in the evaluation of the different difficulty groups according to the judges' opinion (groups were written in increasing order within the indicator). Globally, we can state that there's a statistically relevant difference (Friedman test, $P = 0.000$) in the objectivity for the different difficulty groups.

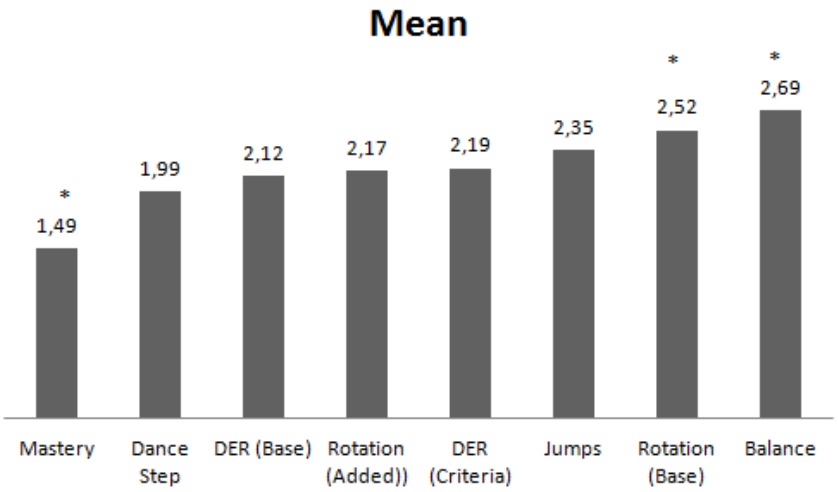


Figure 1. Average indicator of objectivity in evaluation in the different difficulty groups.

Table 3. Sign test between difficulty groups by the objectivity in the evaluation.

	Sign Test						
	Dance Step-Mastery	DER(base)-Dance Step	Rot(Added)-DER(Base)	DER(crit.)-Rot.(Added)	Jump-DER(crit.)	Rot.(Base)-Jump	Balance-Rot.(Base)
Z	-6.904	-1.724	-0.625	-0.256	-0.891	-3.086	-2.585
Asymp.sig	0.000*	0.085	0.532	0.798	0.373	0.002*	0.010*

Comparing groups, we can see in Table 3 that the Mastery show a significant difference (P = 0.000 from all of the others groups, being the Mastery the group where the evaluation is seen by the judges as less objective. Also, between Rotations (base) and Jumps, there are statistically significant differences (*P* = 0.002), being the evaluation less objective in *Jumps* rather than in *Rotations (base)*. The same happens between the groups of *Balance* and *Rotations (base)* (*P* = 0.010), where the evaluation is less objective in *Rotations (base)* than in *Balance*.

When comparing in pairs the groups *DER (base)*, *Rotation (Rot Add)*, *DER (Criteria)* and *Jumps* we can see that there are no statistically significant differences on the degree of objectivity in evaluation. We can even say that the objectivity in evaluation as seen by the judges holds a similar distribution in the four groups (*P* = 0.117).

We can see in Table 4, the summary of the collected data from the judges' opinion about the objectivity in evaluating Execution in Technical and Artistic Faults.

The Technical Faults item is considered the one with higher objectivity in evaluation with the answer "objective to evaluate" getting 80.2% of the answers and the Artistic Faults item is considered the one with less objectivity in evaluation with the answer "Less objective to evaluate" holding 64.8% of the answers. The difference in objectivity in the evaluation of Artistic Faults and Technical Faults is statistically significant (Friedman test, *P* = 0.000).

Table 4. Descriptive statistics of the Execution faults by the objectivity in the evaluation.

Execution Faults		
	Technic Faults	Artistic Faults
Objectivity of Evaluation:	*Frequency (%)*	
Less objective	1.9	**64.8**
More or less objective	17.9	29.6
Objective	**80.2**	5.6
Friedman Test	*P* = 0.000	

Regarding the Artistic Faults sub-items (table 5), the Unity of Composition item is considered the less objective one in evaluation with the answer “less objective to evaluate” collecting 47.5% of the answers. Next, in ascending order of objectivity, the items Music/Movement, and Body Expression were considered “more or less objective to evaluate” (45.7% and 56.6%, respectively). The higher values of objectivity in evaluation are seen in the item Use of Space with the answer “Objective to evaluate” getting 54.3%. Globally, there are statistically significant differences (Friedman, *P* = 0.00) in the objectivity of the evaluation of the different items within the Artistic Faults group, being the degree of objectivity higher in some items than others.

Table 5. Descriptive statistics of the Artistic Faults by the objectivity in the evaluation.

	Artistic Faults			
	Unity Composition	Music/ Movement	Body Expression	Use Space
Objectivity of Evaluation:	*Frequency (%)*			
Less objective	**47.5**	37.0	16.7	23.5
More or less objective	42.0	45.7	**56.8**	22.2
Objective	10.5	17.3	26.5	**54.3**

The Artistic Faults item is considered the one with less objectivity in evaluation with the answer “Less objective to evaluate” holding 64.8% of the answers. Regarding the Artistic Faults sub-items, the Unity of Composition item is considered the less objective one in evaluation with the answer “less objective to evaluate” collecting 47.5% of the answers. Next, in ascending order of objectivity, the items Music/Movement, and Body Expression were considered “more or less objective to evaluate” (45.7% and 56.6%, respectively). The higher values of objectivity in evaluation are seen in the item Use of Space with the answer “Objective to evaluate” getting 54.3%. Globally, there are statistically significant differences (Friedman, *P* = 0.00) in the objectivity of the evaluation of the different items within the Artistic Faults group, being the degree of objectivity higher in some items than others.

According to the judges’ opinion, we can see in Figure 2 the average indicator of objectivity in the evaluation of Artistic and Technical Faults (groups are in ascending order within the indicator).

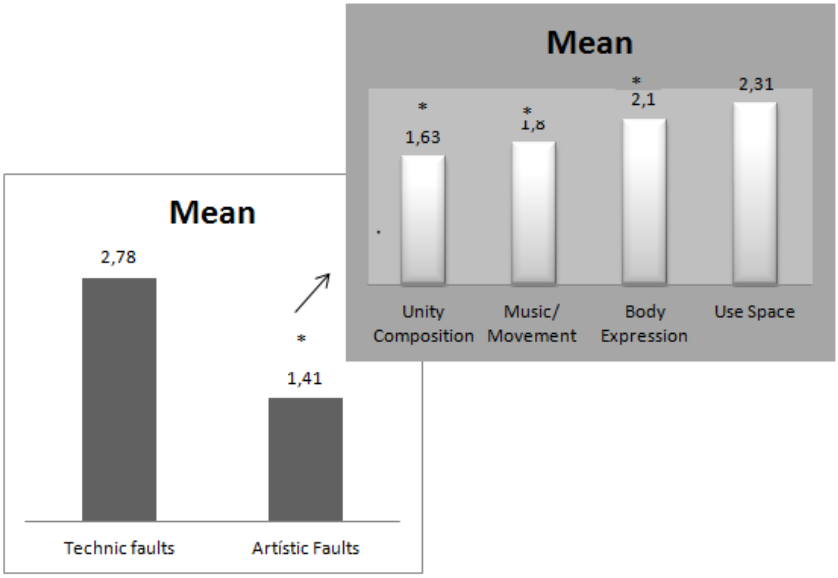


Figure 2. Average indicator of objectivity in the evaluation of Artistic and Technical Faults. Friedman test $P = 0.000^{*}$

Table 6. Sign test between Artistic Components by the objectivity in the evaluation.

	Sign Test		
	Music Movement-Unity Composition	Body Expression-Music Movement	Use Space-Body Expression
Z	-3.125	-4.533	-2.388
Asymp.sig	0.002*	0.000*	0.017*

When we try to analyze if there are significant differences between the items distributions, we see in Table 5 that the item Unity Composition differs from the item Music Movement ($P = 0.002$), with also a significant difference from all the others items, once it's in the item Unity Composition that the evaluation is seen by the judges as less objective. Also between the items Body Expression and Music Movement, there are statistically significant differences ($P = 0.000$), being the evaluation less objective in Music Movement than in Body Expression. The same happens between the items Use of Space and Body Expression ($P = 0.017$), where the evaluation is less objective for Body Expression than for Use of Space.

Concerning the Difficulty, in the opinion of the judges, the evaluation criteria for Mastery should be changed, holding 69.8% of the answers (Figure 3). The Balances group is the one that should not suffer any alteration in the evaluation criteria, holding 3.1% of the answers.

For the Execution, the Artistic Faults evaluation criteria should be changed (63.6%), in opposition to the Technical Faults, which got only 3.7% of the answers.

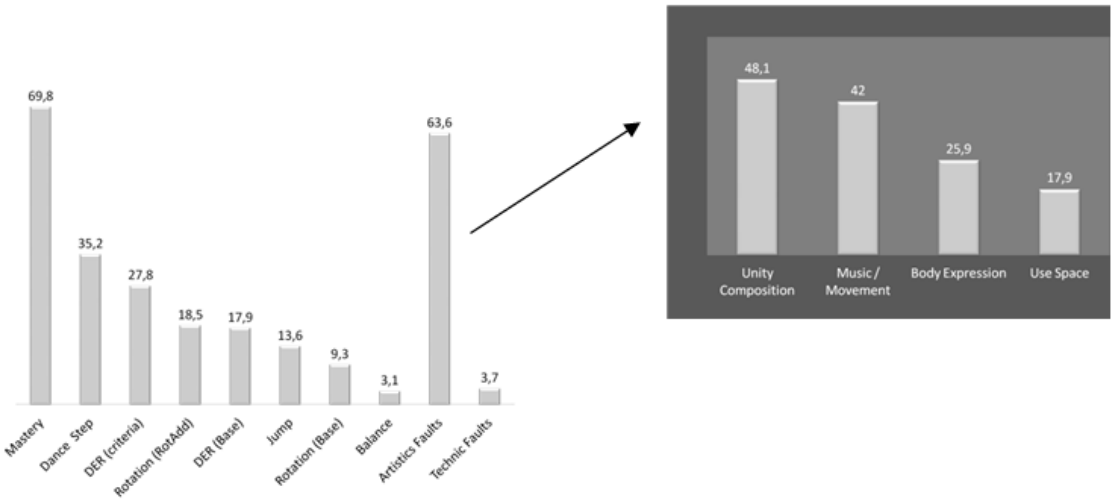


Figure 3. Frequency table of the difficulty and execution groups by the judges opinion be changed.

Studing the items of the Artistic Faults, the evaluation criteria for the Unity of Composition and Music/Movement should be changed, holding 48.1% and 42% of the answers, respectively. The item Use of Space is the one that should not suffer any alteration in the evaluation criteria, holding 17.9% of the answers.

We can see in Table 7 that the majority of the judges (64.8%) agree that there should be a limit of repetitions for the body difficulty in the different apparatus, to improve the variety and the composition of RG routines.

Table 7. Frequency of the judges' opinion ("agree that there should be a limit of repetitions for the body difficulty in the different apparatus").

Frequency Table		
	Frequency	%
NOT	57	35.2
YES	105	64.8
Total	162	100.0

2.2.5. Discussion

The sportive judgement, as human, individual and collective action, holds intelectual, volitional and ethical components, which should be taken into consideration in an integrated global way, so the evaluation of sport performance is done in a responsible manner [6].

In this study, we can verify that the majority of international judges has a high academic level, works as a RG coach and has a large experience in judgement. This type of background offer efficiency conditions to the judges, since they register quality judging values, when compared to quality values of judgement got by judges with less experience, as well as lower academic level [7].

The criteria that can distinguish a "specialist in the matter" goes from the connection of the specialist with the problem, the professional experience, the personal qualities or professional ability, to the guaranty of the quality of the answers and the skill of recognizing detailed information [8].

The most significative difficulties in judgement, come from the excessive amounts of information that the judge needs to quickly summarize in order to transform into a score; in a practical level the judges are able to solve their problems thanks to their experience and knowledge about gymnastics [9], which also allows us to think that the results support the idea that RG judgment is done by individuals who, besides their knowledge of the scoring code, have other resources such as experience and insight, which could be a *plus* in the judgement once the experiences and global vision of the sport lead the judges to deduce, from logic, some aspects that maybe hard to identify or differentiate by younger or inexperienced judges.

Besides de characterization of the judges, we tried to know what they think about the CoP, once they are the ones who use it as an evaluation tool in RG routines. When considered the indicator "objectivity in evaluation" in the different difficulty groups under analysis, the results showed that globally there are statistically significant differences in the objectivity of the evaluation, being the degree of objectivity superior in some groups compared to others. The group with less objectivity in evaluation is Mastery. Following in increasing order of objectivity are Dance Steps, DER (Basis), Rotations (Additional rotations) and DER (criteria).

These results may suggest that the judges find more difficult to evaluate with precision some elements performed by the gymnasts, since their opinion about the evaluation of these groups is less objective probably due to the way the evaluation criteria is described in the CoP, allowing different interpretations.

The evaluation of human performance for some sports is not possible to be done through mechanical ways [10], thus making the reference of the pattern criteria the way to assure validity and reliability in the result of the evaluation, when trying to evaluate the quality of a movement.

It is also important to state that the complexity in the evaluation of the referred difficulty groups (Mastery, Dance Steps and DER) may be also related to the fact that there is no pattern reference in the CoP, compared to what happens for other difficulty groups such as Jumps, Balances and Rotations.

The difficulty groups in which the evaluation is more objective are Jumps, Rotations (Basis) and Balances. These results indicate that in the opinion of the judges, the evaluation criteria of these groups described in CoP allow an objective evaluation, with easy application. It is important to state that the CoP holds a list with pattern images for these three groups, which in our opinion allow an immediate perception of the correctly performed difficulty and consequently the objectivity in evaluation. The criteria to determine the quality of the evaluation should refer to a pattern, model or arbitrary level of that same quality [11].

The study about the degree of agreement between the 4 judges in the evaluation of the different difficulty groups declared in the competition cards used in KIEV 2013 WC, supports the results obtained here [12], once this study highlights the same difficulty groups, where we see more disagreement in the evaluation done by the judges.

When considering the indicator “objectivity in evaluation” within the Execution items, the results showed that globally that there are significant differences in the evaluation objectivity of Technical Faults and Artistic Faults. The Artistic Faults group is considered the less objective in the evaluation. About this group, we found out that there are statistically significant differences in the objectivity of evaluation of the different items that integrate it, with higher objectivity for some items rather than others. The parameter Unity of Composition is considered the less objective in the

evaluation. Following in increasing order of objectivity there are Music/Movement and Body Expression. It's the Use of Space item where we get higher values of objectivity. In the same way, Bučar et al. [13] found results of low validity and reliability in the judgment of artistic components in Artistic Gymnastics, what allows us to consider that the current instruments of evaluation for gymnastics artistic components require monitoring for a possible reassesment and eventual restructuring.

We also tried to identify which evaluation criteria would the judges like to modify in order to potentiate their performances as evaluators. The results found show proposals of changes in the evaluation criteria for Mastery group and for the Artistic Faults group, in particular for "Unity Composition" and "Music/Movement" parameters. Finally, the results indicate that the majority of the judges (64.8%), consider that the CoP should limit the repetitions of body difficulties in the different apparatus, helping to enrich the compositions of RG routines and consequently the evolution of the sport. [14], analyzed the evolution of scoring codes in RG and found out that the evolutive tendency should contemplate variety and diversity allowing exploration of new skills.

2.2.6. Conclusions

The evaluation system to determine the final scores of a RG exercise is given by the CoP, being this an instrument elaborated by the FIG. Yet, the judges are the ones using it as an evaluation tool thus their opinions represent a reference value to be considered in its elaboration. They manifested different opinions about the objectivity in the evaluation of the Difficulty, Execution, as well as the different parameters of evaluation in artistic faults.

They highlighted the Mastery, Dance Steps and DER, in Difficulty and the Artistic Faults (mainly "Unity Composition" and "Music/Movement") in execution, as being the ones with most complexity in evaluation when considered the objectivity. They suggested changes in the evaluation criteria of these groups, in order to become more precise in the final evaluation.

Finally the judges stated that they would like to have in the CoP some rules to limit the repetition of difficulties in the different apparatus routines, in order to improve the diversity and variety in RG routines, promoting an enrichment of the sport.

Therefore, we conclude that the instrument of evaluation used right now is not yet ideal to absolutely assure the validity and reliability in RG judgment. These results may contribute for a reconstruction of the CoP and consequently help in the evolution of the sport.

We expect that new rules of artistry evaluation will bring improvement of reliability and consistency of judges and this should be verified through further research of future competitions.

Acknowledgments

We are especially grateful to International Gymnastics Federation (FIG). We also thank to international Rhythmic Gymnastics Judge.

2.2.7. References

[1] FIG-International Gymnastics Federation. 2012. "Code of Points for Rhythmic Gymnastics Competitions, 2012-2016." Accessed November 18, 2012. http://www.fig-gymnastics.com/site/page/view?id=472

[2] Flessas, K., and Mylonas, G. 2015. "Judging the Judges' Performance in Rhythmic Gymnastics." *Medicine & Science in Sports & Exercise. American College of Sports Medicine* 47 (3): 640-8.

[3] Guardo, M. 2004. "Toward a Theory of Arbitration for Sport." *Rev. Digital. Buenos Aires* 68. Accessed July 8, 2004. http:/www.efdeportes.com/ (in Spanish)

[4] Martin, S. 2006. "Desirable Characteristics for Arbitration in Sports Trial in Judo". *Latin American Journal on Exercise and Sport Psychology* 1: 27-40. (in Spanish)

[5] Fernandez-Villarino, M. 2013. "Practical Skills of Rhythmic Gymnastics Judge." *Journal of Human Kinetics* 39/2013: 243-249.

[6] Palomero, M. L. 1996. "Towards an Objectification of the International Code of Points Rhythmic Gymnastics." Ph.D. thesis, The Barcelona University. (in Spanish).

[7] Leandro, C., Ávila-Carvalho, L., and Lebre, E. 2010. "The Avaluation of the Performance of Rhythmic Gymnastics' Judges." *Palestrica of the Third Millennium Civilization & Sport* 11 (3): 202-6.

[8] Almenara, J. 2013. “The Use of Expert Judgment for the Evaluation of TIC: The Coefficient of Expert Competence.” Ph.D. thesis, Sevilha University. (in Spanish).

[9] Plessner, H. 2005. “Positive and Negative Effects of Prior knowledge on Referee Decisions in Sports.” In *The Routines of Decision Making*, edited by Betsch, T., and Haberstroh, S. 311-24.

[10] Morrow, J., Jackson, A., G-Disch, J., and Mood, D. 1995. *Measurement and Evaluation in Human Performance.* Library of Congress Mystat, USA.

[11] Simões, G. 2000. *The Evaluation of Teacher Performance.* Texto Editora, Lisboa, Portugal. (in Portuguese).

[12] Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., and Bobo-Arce, M. 2015. “Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines.” *Science of Gymnastics Journal* 7 (3): 81-93.

[13] Bučar Pajek, M., Kovač, M., Pajek, J., and Leskošek, B. 2014. “The Judging of Artistry Components in Female Gymnastics: A Cause for Concern?” *Journal of Human Kinetics* 37/2014: 173-81.

[14] Liu, X. X., and Kuang, L. 2001. “Review of Evolvement Course of International Evaluation Rules in Rhythmic Gymnastics and Its Effects on Technique Development.” *Journal of Beijing Sport University* 3 (24): 412-5.

2.3. Study III: Judging in Rhythmic Gymnastics at different levels of performance

Judging in Rhythmic Gymnastics at different levels of performance.
Leandro, C. [1,3], Ávila-Carvalho, L. [2], Sierra-Palmeiro, E. [3], Bobo-Arce, M. [3]
Human kinetics Journal. (Submited)

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

2.3.1. Abstract

This study aims to analyse the quality of Difficulty judging in Rhythmic Gymnastics, at different levels of performance. The sample consisted of 1152 Difficulty scores concerning 288 individual routines, performed in the World Championship 2013. The data were analysed using the mean absolute judge deviation from final difficulty score, Cronbach's alpha coefficient and intra-class correlations, for consistency and reliability assessment. For validity assessment, mean deviations of judges' Difficulty scores, Kendall´s coefficient of concordance W and ANOVA eta-squared values were calculated. Overall, the results in terms of consistency (Cronbach's alpha mostly above 0.90), and reliability (intra-class correlations for single and average measures above 0.70 and 0.90 respectively) were satisfactory, in the first and third parts of the ranking on all apparatus. The medium level gymnasts, those in the second part of the ranking, have inferior reliability indices and highest score dispersion. In this part, the minimum of corrected item-total correlation of individual judges was 0.55, with most values well below this, and the matrix for between-judge correlations identified remarkable inferior correlations. These findings suggest that the quality of Difficulty judging in Rhythmic Gymnastics may be compromised at certain levels of performance. In future, special attention should be paid to the judging analysis of the medium level gymnasts, as well as the Code of Points applicability in this level.

Key words: Rhythmic Gymnastics; evaluation; bias, validity; reliability.

2.3.2. Introduction

In artistic sports like Rhythmic Gymnastics (RG), the performance in competition is evaluated by judges that apply a tool (Code of Points) and give a score that determines the value of the routine and the position of the gymnast in a final ranking. Since the performance does not come out from an objective measure but from a complex judging process, quite often RG is considered to be a subjective sport (Gateva, 2014).

Recent research has paid attention mainly to the experience and the capacity of the judges to use cognitive and perceptual strategies to interpret and register gymnast's performance in competition (St. Marie et al., 2001; Plessner and Schallies, 2005; Dallas and Kirialanis, 2010; Heinen et al., 2012). Furthermore, research has also emphasized the judges' need for developing a set of skills that contributes to an effective assessment process (Fernandez-Villarino et al., 2013), and to the overall error detection efficiency (Flessas et al, 2015). In a RG competition the performance is evaluated by two panels of judges: the Difficulty (D) jury which judges the routines' content (what the gymnast performs) and the Execution (E) jury which evaluates the quality of the routines (how the gymnast performs). The present Code of Points states that minimum four judges are required on the D jury, as well as on the E jury. For both, the final score is determined calculating the average of two intermediate scores (FIG, 2012).

The judging process for Difficulty and Execution evaluation is different. Difficulty judges have to check the content of the routines that is stated and signed by the coaches in the specific forms. Their task is to validate the difficulty elements declared while the gymnast performs her routine. These difficulty elements may range from 0.1 to 1.5 points or more, up to a total maximum of 10 points. Preciseness in the judgement is needed since differences between the judges may cause great deviations in the final D score, and this score has a great influence in the gymnast's final position in the ranking (Cuk et al., 2012; Leskosek et al., 2015).

In higher level competitions, the more experienced judges are assigned to evaluate the difficulty component of the routine. However, quite often the differences between the athletes performance are so small, that little and consistent mistakes made by the judges may interfere in the final classification of the gymnast (Bucar et al., 2011; Bucar et al., 2013).

Consequently, to verify the quality of judging in rhythmic gymnastics it is needed to identify the extent to which the scoring system is objective. Therefore reliability and validity of the scores must be proof.

The aim of this study is to analyze the reliability and validity of the difficulty scores of individual routines in RG at different levels of performance and with different apparatus. It was hypothesized that (i) the level of performance of the gymnast may affect the reliability and validity of the scores in competition and (ii) the type of apparatus used (hoop, ball, clubs and ribbon) may affect the reliability and validity of the scores in competition.

2.3.3. Methods

Participants

The sample consisted of 1152 Difficulty scores corresponding to 288 exercises performed at the Kiev World Championships 2013, clustered according to the position of the gymnast on the final ranking (1st part, 2nd part and 3rd part) and to the apparatus (hoop, ball, clubs and ribbon). The scores were obtained from the official book of results of the qualification competition. The study was ethically approved by the International Gymnastics Federation. Full blinding of the judges involved was undertaken. To protect the judges' anonymity we randomly changed their position in the analysis from the book of results.

Procedure

The sample was divided into three groups according to the gymnast's final rankings in each apparatus: the first part of the ranking (top 24 gymnasts) the second part of the ranking (medium 24 gymnasts) and third part of the ranking (last 24 gymnasts), to allow the comparison of the reliability and validity values at different performance levels with the all 4 apparatus. For each of the groups, four judges' D scores were considered.

Statistical analysis

For each group (top, medium and last gymnasts) descriptive statistics for D score was calculated, as well the distributional statistics (mean and standard deviation) for individual judge's D score and mean deviation from final D score. This mean deviation is a measure of bias (systematic under- or over-estimation) and provides information related to the validity of scoring. Segundo (Bučar et al., 2011), when examining validity, the ideal test of validity would have to implement a comparison of concrete judging with the gold standard of judging performance; however no such gold standard currently exists. It is possible however, to focus on a special case of validity, which deals with the presence of systematic over or under - rating or scoring of competitors - what is also called bias.

Additionally, two analyses of between-judges differences were performed: Kendall's concordance coefficient and ANOVA repeated measures to identify possible systematic bias.

The correlation between individual judge's scores and total scores was also calculated.

The consistency and reliability assessment of the evaluation was measured using the Cronbach's alpha coefficient for each group of judges on each apparatus. Two types of intra-class correlation (ICC) were calculated: the single measure ICC and the average measures ICC.

Data were analyzed using the Statistical Package for the Social Sciences - Version 21.0 (SPSS 21.0, Chicago, USA) and Microsoft Office Excel 2010.

2.3.4. Results

The variability of D scores (dispersion) is in general larger for the 2nd part of the ranking and is relatively smaller in the 1st part of the ranking, in hoop, ball, clubs and ribbon, (table 1). The average value of the D score for the different apparatus does not show great variability in each part of the ranking. The worst individual deviations in judging for each part of the ranking and apparatus (all remaining individual judge values were better) are presented. Besides the worst deviations, also the smallest values for item-total correlation are indicated as well as the Cronbach's alpha coefficient for each apparatus. It can be seen that maximal individual judge mean deviations from the final

D score are overall relatively small, all of them below 0.2. Only in the 3rd part of the ranking in hoop and clubs we find maximum deviations with values 0.38 and 0.33 respectively. In terms of measures of common performance for the 2nd part of the ranking in all apparatus we obtain the poorest values of Cronbach's alpha and the smallest values of minimum item-total correlation. However, most of the values are still above 0.9 in the 1st and 3rd part of the ranking. We can't see great differences between the different apparatus.

Table 1. Statistics of D scores and the performance individual Judge.

	Apparatus	Mean±SD	Dev. Máx.	Ab.Dev.Máx	R min.	Cα
1st Part of the Ranking	Hoop	8.25±0.53	-0.15	0.29	0.71	0.91
	Ball	8.34±0.49	-0.03	0.20	0.76	0.91
	Ribbon	7.98±0.60	-0.12	0.26	0.75	0.92
	Clubs	8.21±0.55	-0.20	0.29	0.76	0.91
2nd Part of the Ranking	Hoop	6.61±0.45	-0.13	0.41	0.16	0.65
	Ball	6.85±0.60	0.26	0.41	0.47	0.79
	Ribbon	6.52±0.63	-0.08	0.38	0.55	0.77
	Clubs	6.68±0.48	-0.16	0.42	0.24	0.59
3rd Part of the Ranking	Hoop	4.58±1.31	0.38	0.56	0.82	0.94
	Ball	4.64±1.33	-0.22	0.36	0.89	0.96
	Ribbon	4.36±1.38	-0.10	0.32	0.92	0.97
	Clubs	4.56±1.39	0.33	0.44	0.88	0.95

Minimum (min) and maximum (max) values, mean and standard deviation (SD), Dev max: maximal judge average deviation from D score, Ab dev max: maximum of average absolute deviation from D score; R min: minimum of corrected item-total correlation of individual judges; Cα: Cronbach's alpha coefficient.

When testing the inter-judge differences with repeated measures ANOVA, eta-squared values representing the bias effect size - Figure 1. These values are quite concordant with Kendall`s results. Kendall's W is statistically significant in the 1st part of the ranking for Hoop and Clubs, in the 2nd part of the ranking for the Ball and in the 3rd part of the ranking for the Hoop.

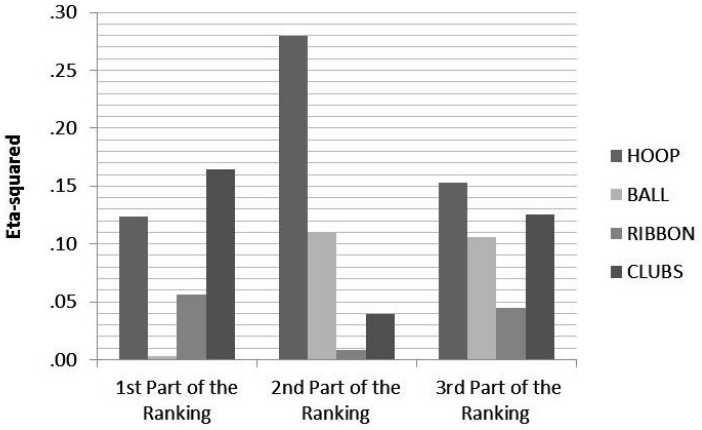


Figure 1. The eta-squared (η2) values of repeated measures ANOVA of D-scores in all apparatus clustered according to the position of the gymnast on the final ranking (1st part, 2nd part and 3rd part)

For the analysis of between-judge correlations; the Pearson's correlation coefficients are shown in table 2. It can be seen that most of the correlation coefficients are above 0.7 in the 1st and 3rd parts of the ranking. In the 2nd part of the ranking a high number of correlations are below 0.5.

Table 2. Correlation Matrix for between - judges correlation.

		1st part of the ranking			2nd part of the ranking			3rd part of the ranking		
	Judge	2	3	4	2	3	4	2	3	4
	1	0.72**	0.76**	0.85**	0.42*	0.15	0.73**	0.86**	0.87**	0.75**
Hoop	2		0.48*	0.80*		0.33	0.30		0.85**	0.81**
	3			0.74**			-0.14			0.78**
	1	0.71**	0.80**	0.68**	0.71**	0.49*	0.47*	0.86**	0.83**	0.89**
Ball	2		0.74**	0.79**		0.34	0.55**		0.89**	0.79**
	3			0.61**			0.38			0.86**
	1	0.64**	0.79**	0.69**	0.38**	0.44*	0.40*	0.90**	0.91**	0.93**
Ribbon	2		0.74**	0.82**		0.45*	0.32		0.87**	0.93**
	3			0.87**			0.61**			0.92**
	1	0.74**	0.83**	0.69**	0.23	0.19	0.12	0.83**	0.83**	0.86**
Clubs	2		0.70**	0.71**		0.34	0.36		0.88**	0.83**
	3			0.69**			0.34			0.84**

*** correlation is significant at the 0.01 level (2 - tailed);*

** correlation is significant at the 0.05 level (2 - tailed)*

Overall measures of inter-judge reliability are shown in table 3. The poor concordance of judges on 2nd part of the ranking (as evident from Cronbach's alpha coefficients), can be inferred also from the calculated ICC of single values, otherwise the observed ICC values are high, mostly above 0.7. The values of ICC for average values are

quite close to Cronbach's alpha coefficient values. In the second part of the ranking, the ICC for single values confirmed the highest sensitivity for the deviations in inter-judge agreement and reliability if compared to other measures (Cronbach's alpha and ICC for average measures).

Table 3. Overall measures of inter-judge reliability.

	Apparatus	ICC Single	ICC Average	Kendall´s W	P (W)
1st Part of the Ranking	Hoop	0.70	0.90	0.133	0.023*
	Ball	0.73	0.91	0.013	0.821
	Ribbon	0.76	0.92	0.053	0.283
	Clubs	0.88	0.89	0.177	0.005*
2nd Part of the Ranking	Hoop	**0.33**	0.66	0.021	0.671
	Ball	**0.46**	0.77	0.109	0.049*
	Ribbon	**0.47**	0.78	0.002	0.982
	Clubs	**0.26**	0.58	0.071	0.162
3rd Part of the Ranking	Hoop	0.80	0.94	0.112	0.045*
	Ball	0.84	0.95	0.088	0.096
	Ribbon	0.91	0.97	0.028	0.564
	Clubs	0.83	0.95	0.064	0.203

ICC single (average): intra-class correlation for single (average) scores; p(w): p value of Kendall´s W

2.3.5. Discussion

The aim of the current study was to analyse the quality of Difficulty judging on Rhythmic Gymnastics at different levels of performance. To the best of our knowledge, this is the first study that has analysed the reliability and validity of Difficulty judging, perhaps, because only in this Olympic cycle is the final score determined calculating the average of two intermediate scores (of 4 judges) and not for consensus, a joint score of two judges.

Overall, the results suggest that the reliability of the judgment in RG is satisfactory in the first and third parts of the ranking, once the Cronbach's alpha is above 0.90, minima of item total correlations and the ICC of average scores are above 0.80. Overall, for the World Championships analyzed, regarding the final ranking of the gymnasts, the indices of consistency are satisfactory in both high and low level gymnasts.

The level of the consistency indices was lower in the 2nd part of the ranking. When trying to explain the inferior reliability results for medium level gymnasts it is valuable to inspect the between-judge correlation matrix, as many of the reliability measures of

judges' performances are based on Pearson's correlations. We can identify several judges (without highlighting any over the others) whose correlation coefficients are below 0.5 in all apparatus.

The validity in our analysis was assessed through systematic bias in judging, considered as repetitive under- or over-estimation of particular judges. When looking at the results as a whole, systematic bias in individual judge's scores and judges' panels was modest or poor in the 2nd part of the ranking. Popovic (2000) also detected international bias in judging rhythmic gymnastics at the Sydney 2000 Olympic Games. It is obvious that the quality of judging differs when evaluating different levels of gymnasts' performances. There are numerous objective and subjective factors for those differences. To Ferreirinha and Carvalho (2012) besides these external factors that may lead the judges to commit mistakes that go further away from the dimension of conscience and therefore are not intentional, there are other factors related directly to the evaluation rules (code of points) that may be in the origin of these deviations. Also Bucar et al. (2014) found out similar results when analysed the evaluation of the artistic component in female gymnastics. Fernandez-Villarino et al. (2013) claims that the specific situation in which the judges must evaluate during the same competition gymnasts of different ages and different levels may create problems in the ability to discriminate performances.

Our results show that the biases in the judgment of rhythmic gymnastics competition routines are not so much due to the performance of specific judges, but more to the differences in the level of performance of the gymnasts at the same competition.

To further clarify the factors contributing to the observed, we can speculate that these differences are perhaps a source of additional variability in the judge's scores and that part of the problem may originate in the judging rules (Code of Points) that are not well defined to evaluate the gymnasts whose execution is not clear and precise. This situation according to Debien et al. (2014) may be a source of variability between the judges, caused by stress, which appears due to the acknowledgment of something which is not expectable. The apparatus used by the gymnast does not seem to be a cause of variability in the judging since we found equivalent values of mean deviations from final D scores for all apparatus, in each of the parts of the ranking

In conclusion, the comparison of reliability and validity indices brought attention to medium level gymnast, which seems more vulnerable to deviations from high reliability indices found in other levels gymnasts (1st and 3rd parts of the ranking). Further work must be done to explain the inferior results at medium level gymnastics and test the solutions for improvement. This study provides updated information about the individual routines judgment in rhythmic gymnastics, to be considered for possible modifications of the present Code of Points, in particular for the definition of the evaluation criteria in order to reach higher levels of reliability and validity in judgment.

2.3.6. References

Bučar M, Cuk I, Pajek J, Karacsony I, Leskosek B. Reliability and validity of judging in women's artistic gymnastics at University Games 2009. Eur J Sport Science, 2011; 12: 207-215

Bučar M, Cuk I, Pajek J, Kovac M, Leskosek B. Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? *J Hum kinet,* 2013; 37: 173-181

Bučar M, Kovač M, Pajek J, Leskošek B. The Judging of artistry components in female gymnastics: a cause for concern? *Sci Gymnastics J*, 2014; 6 (3): 5-12

Catteeuw P, Helsen W, Gilis B, Wagemans J. Decision-making skills, role specificity, and deliberate practice in association football refereeing. *J Sport Sci*, 2009; 27: 1125-1136

Čuk I, Fink H, Leskošek B. Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. *Sci Gymnastics J*, 2012; 4: 73- 82

Dallas G, Kirialanis P. Judges' evaluation of routines in men's artistic gymnastics. *Sci Gymnastics J*, 2010; 2: 49- 58

Debien P, Noce F, Debien J, Costa V. Stress in rhythmic gymnastics refereeing: a systematic review. *Revista da Educação Física/UEM*, 2014; 25 (3): 489-500

Ferreirinha J, Carvalho J.Tendencies and deviations of judging in gymnastics.*Revista ENGym*, 2012; 1: 2-3

Fernandez-Villarino M, Bobo-Arce M, Sierra-Palmeiro E. Practical Skills of Rhythmic Gymnastics Judges. *J Hum kinet,* 2013; 39: 243-249

FIG-International Gymnastics Federation. Code of Points for Rhythmic Gymnastics Competitions, 2012; Available at: http://www.fig-gymnastics.com/site/page/view?id=472; accessed on 15.11.2012

FIG- International Gymnastics Federation. Gymnastics Results, 2013; Available at: http/www.gymnasticsresults.com; accessed on 01.10.2013

Flessas K, Mylonas G, Panagiotaropoulou D, Tsopani A, Korda C, Siettos A, Di cagno I, Evdokimidis E, Smyrnis J. Judging the Judges' Performance in Rhythmic Gymnastics. *Med Sci Sport Exer . American College of Sports Medicine,* 2015; 47(3): 640-648

Gateva M. Investigation of the effect of the training load on the athletes in rhythmic and aesthetic group gymnastics during the preparation period, *Research in Kinesiology,* 2014; 4: 40-44

Heinen T, Vinken P, Velentzas K. Judging Performance In Gymnastics: A matter of motor or visual experience? *Sci Gymnastics J,* 2012; 4: 63- 72

Leskošek B, Čuk I, Bučar-Pajek M. Trends in E and D scores and their influence on final results of male gymnasts at European Championships 2005–2011. *Sci Gymnastics J*, 2015; 5 (1): 29-38

Plessner H, Schallies E. Judging the cross on rings:a matter of achieving shape constancy. *Appl Cognitive Psych*, 2005; 1: 1145-1156

Popovic´ R. International bias detected in judging rhythmic gymnastics competition at Sydney 2000 Olympic Games. *Facta Universitatis* 2000; 1: 1-13

Ste-Marie D M, Valiquette S M, and Taylor G. Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. *Res Q Exercise Sport*, 2001; 72(4): 420-426

2.4. Study IV: Departure Difficulty Score Vs Final Difficulty Score. The effect of Performance in Elite Rhythmic Gymnastics

Departure Difficulty Score Vs Final Difficulty Score. The effect of Performance in Elite Rhythmic Gymnastics. Leandro, C. [1,3], Ávila-Carvalho, L. [2], Sierra-Palmeiro, E. [3], Bobo-Arce, M. [3]
Athens Journal of Sports. 2016. *Vol. X, No. Y*

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

2.4.1. Abstract

The aim of this study was to determine how the difference between the departure difficulty score, and the final difficulty score, in the different type of elements affects the performance success in elite rhythmic gymnastics. Two hundred eighty-eight individual routines performed at the World Championship 2013 were analysed. Non-parametric test- Kruskal-Wallis test, was applied to determine whether there were significant differences between different level groups. The results show significant differences between the DDS and FDS, which increase as the level of performance decreases. Rotations, Dynamic Elements with Rotation and throw and Mastery are the main types of difficulty elements responsible for this difference. These results suggest that the judges and coaches do not have the same perception of the evaluation criteria of the difficulty elements. The findings can contribute to improve the definition criteria of the difficulty elements and to clarify the specific needs of the training program.

Key Words: Difficulty score, coach, judge, rhythmic gymnastics.

2.4.2. Introduction

The ability to control and monitor the technical content of competitive individual routines in Rhythmic Gymnastics accurately is an important factor of effective high performance in this sport. The knowledge can also contribute to preview and characterise the effort requirements allowing improvements in gymnasts' preparation for elite competitions (Ferreirinha, 2011). A precise understanding of the technical content of individual RG routine, acknowledged by the judges' evaluation, can be beneficial to the International Gymnastics Federation (Cuk et al., 2012), as well as for coaches and gymnasts that can also be used as feedback to adapt the training structure in order to improve performances (Fernandez-Villarino, 2015).

Several studies (Caburrasi, 2003; Ávila, 2011; Ávila, 2012; Trifunov, 2013; Agopyan, 2014), analysed the number and the level of difficulty elements presented by the coach on the competition form. The number and the level of difficulty elements is the Departure Difficulty Score (DDS). However, those studies did not emphasise information about the contribution of each type of element acknowledged by the judges, which leads to the Final Difficulty Score (FDS).

The elements prescribed before the competition in the forms do not translate the real success in competition, which suggests that the use of these indicators may not be enough to understand the individual adaptations and establish the optimised training models, (Arkaev, 2004). Also as a consequence of the constant and quick evolution of this sport, a permanent upgrade of these kinds of studies is needed. (Caburrasi, 2003; Cuk, 2012; Massidda, 2012; Hokelmann et al., 2012; Bucar, 2013, Pelin, 2013).

An individual RG routine is composed of body and apparatus elements linked in a specific way, which are called difficulty elements. The code of points (CoP) holds a great variety of difficulty elements to be used in the routines. One important characteristic of the RG is to allow the gymnast to link it in her own way, with a stylish presentation, clever configuration, and prefect presentation (Wang, 2013). On the present Olympic cycle, the content of individual RG routine should respect the specific requirements that are common to the routines of all 4 apparatus: Jumps/Leaps, Balances, Rotations, Apparatus Mastery, Dance Steps and Dynamic Elements with Rotation and throw (DER) (FIG, 2012). The value of each difficulty element is from 0.10 points to 1.50 points or more, which may be absolutely determinant in the final score obtained in competition.

The inclusion of complex abilities in the routines is essential to have a high score (Massida, 2012).

It is important to understand which group(s) of elements contribute the most for the difference between the DDS proposed by the coach on the difficulty form, and the FDS given by the judge in competition. Once the coaches know the cause of this difference, they may optimise the training process on these groups of elements, promoting an effective success in competition. Thus, before suggesting eventual future changes, it is important to understand how , find out what should be changed and what should be maintained.

The aim of this study is to determine the real performance success in elite rhythmic gymnastics through the analyse of the difference between the departure difficulty score and the final difficulty score. The analysis will be done (i) according to the final ranking of the gymnast in competition in order to see if the technical level of the gymnasts has influence in the results, and (ii), according to the type of element, to determine if there are elements with more influence in the difference between the DDS and the FDS.

2.4.3. Methods

Participants

One thousand and one hundred and fifty-two difficulty forms concerning 288 individual routines were analysed (4 forms per routine, 1 per judge). The routines were performed by gymnasts from 45 different countries competing at Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013.

Measures/ Procedure

All difficulty elements reported in the difficulty forms provided by the gymnasts at the competition were recorded. The evaluation of the each difficulty element was considered according to the average of the 2 intermediate scores done by the 4 judges on the form. The analysis was done considering the sample clustered into 3 groups according to gymnasts final ranking as follows: The top 24 gymnasts on the ranking (Group 1), the 24 middle gymnasts on the ranking (Group 2) and the 24 lower placed gymnasts on the ranking **(**Group 3**)**. This division allowed for the comparison of the routine difficulty value declared by the coach in the difficulty form (DDS) with the

difficulty score given by the judges during the competition (FDS), for gymnasts with different technical levels. Then, we studied the sample according to the type of difficulty element performed, listed according to the composition requirements of the Code of Points (FIG, 2012).

Full blinding of the judges and gymnasts involved was undertaken. That is, in order to protect the judges' and gymnasts' anonymity we blinded their names.

The forms were analysed by two international RG judges. The intraclass correlation coefficient (ICC) in test-retest method (intra-examiner) was 0.99. The ICC between the observers (inter-examiner) was 0.98.

Statistical analysis

The data were analysed using the Statistical Package for Social Sciences – version 20.0 (SPSS 20.0, Chicago, USA) and Microsoft Office Excel 2007. The level of significance was set at $\alpha = 0.05$ (confidence interval of 95%). Descriptive statistics were calculated using the mean values as a measure of central tendency, standard deviation (SD) as a measure of dispersion. After checking the normalities in the data distribution ($p<0.05$) using the Kolmogorov-Smirnov normality test, we resorted to a non-parametric test-Kruskal-Wallis test, to determine whether there were significant differences between the three groups in the Rhythmic Gymnastics World Championship ranking. A multiple regression was used to analyse the influence of each difficulty element in the gymnasts' final difficulty score.

2.4.4. Results

Comparing the routine difficulty value declared by the coach in the difficulty form (DDS) with the difficulty score given by the judges during the competition (FDS) we get the results summarized in Table 1 where we present the average values for DDS and FDS (mean ± sd) and the difference (Δ) between these values in the 3 groups of the gymnasts' final ranking.

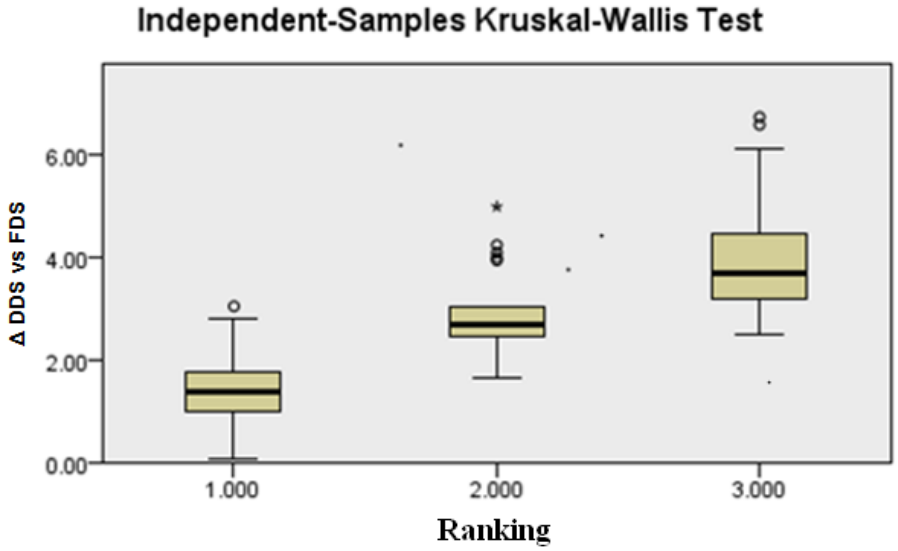
Table 1. Descriptive values of Departure Difficulty Score and Final Difficulty Score.

Routine Difficulty Score	**Group 1 (n=96)**	**Group 2 (n=96)**	**Group 3 (n=96)**
	Mean±sd	**Mean±sd**	**Mean±sd**
DDS	9,83±0,22	9,59±0,26	8,50±0,98
FDS	8,42±0,64	6,81±0,62	4,61±1,37
Δ DDS vs FDS	1,41±0,58*	2,78±0,57*	3,88±0,94*

*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score; Δ DDS vs FDS: differences between DDS and FDS

We could see, by only observing the numbers, that there is a considerable difference between the DDS and the FDS, even in the 24 best gymnasts (1.41±0.58 points). This difference is almost twice as higher in group 2, the 24 middle gymnasts (2.78±0.57 points) and for group 3, it achieves an average of 3.88±0.94 points. We found statistically significant differences between these results, visible in the Figure 1.

Figure 1. Boxplot for difference between DDS and FDS for the 3 parts of the gymnasts' final ranking.



*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score; Δ DDS vs FDS: differences between DDS and FDS

After getting these results we analysed the differences between de DDS and FDS, considering the different types of difficulty elements, trying to find out if the differences are mainly connected to some type of difficulty elements, and if the type of elements with higher difference are the same for the 3 groups of the gymnasts ranking. In table 2 we can observe the DDS and FDS for each type of difficulty element for the gymnast in the different groups.

Table 2. Descriptive values of the different types of the difficulty elements of Departure Difficulty Score and Final Difficulty Score.

Elements Difficulty Score		**Group 1** **Mean+sd**	**Group 2** **Mean+sd**	**Group 3** **Mean+sd**
Jumps	DDS	1,54±0,44	1,40±0,45	1,25±0,33
	FDS	1,52±0,44	1,23±0,48	0,72±0,49
	Δ DDS vs FDS	0,01±0,07*	0,16±0,25*	0,52±0,43*
Balance	DDS	1,09±0,51	1,18±0,45	1,16±0,40
	FDS	1,05±0,52	0,96±0,46	0,69±0,38
	Δ DDS vs FDS	0,04±0,11*	0,21±0,28*	0,47±0,40*
Rotations	DDS	3,00±0,88	2,89±0,66	2,22±0,71
	FDS	2,32±0,77	1,80±0,47	1,08±0,52
	Δ DDS vs FDS	0,67±0,37*	1,08±0,42*	1,13±0,48*
DER	DDS	2,05±0,30	2,01±0,27	1,83±0,35
	FDS	1,64±0,42	1,40±0,38	1,00±0,51
	Δ DDS vs FDS	0,41±0,37*	0,61±0,33*	0,83±0,44*
Mastery	DDS	0,77±0,40	0,82±0,45	0,96±0,71
	FDS	0,56±0,40	0,32±0,27	0,24±0,27
	Δ DDS vs FDS	0,20±0,18*	0,50±0,31*	0,72±0,59*
MixDif	DDS	0,98±0,20	0,90±0,10	0,88±0,05
	FDS	0,93±0,27	0,76±0,25	0,57±0,28
	Δ DDS vs FDS	0,05±0,13*	0,13±0,24*	0,31±0,30*
Dance Steps	DDS	0,81±0,26	0,76±0,23	0,81±0,30
	FDS	0,79±0,26	0,70±0,23	0,69±0,30
	Δ DDS vs FDS	0,01±0,07*	0,05±0,12*	0,12±0,18*

*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score;
Δ DDS vs FDS: differences between DDS and FDS

As we can see in Table 2, the difference between the DDS and FDS is highly correlated with the final ranking of the gymnasts in each Difficulty element. The lowest is the position of the gymnast in the ranking, the highest is the difference between the departure and the final score in all types of difficulty element. we can distinctively see the difference between the DDS and FDS behaviours according to the different types of difficulty elements. For the Jumps, Balances, MixDif and Dance Steps the difference between the DDS and the FDS is very low for the best gymnasts and increases a lot as

we go down in the gymnasts' ranking. For the Rotations, DER and Mastery the difference between the DDS and the FDS is very high in the 3 groups.

The Figure 2 shows the contribution of each type of element (in percentage) to the difference between the DDS and the FDS according to the final ranking of the gymnast.

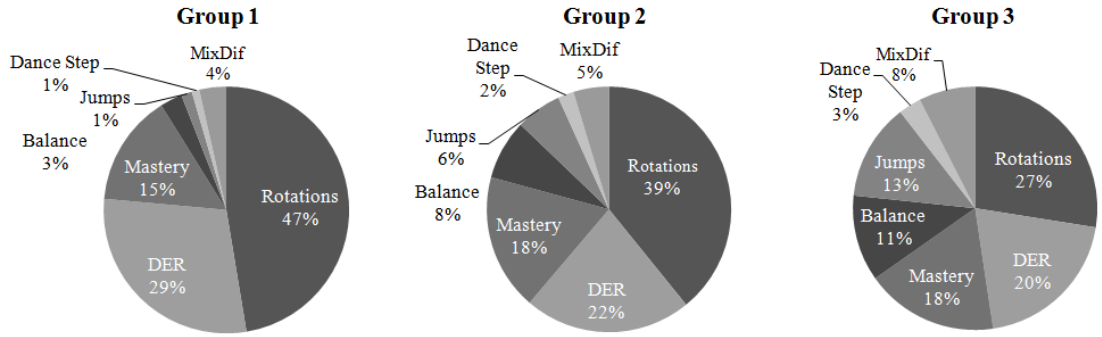


Figure 2. Contribution of each type of element (in percentage) to the difference between the DDS and the FDS, for the 3 parts of the gymnasts' final ranking

We can see that the Rotations clearly contribute, in the 3 groups of the ranking (47%, 39% and 27%) to the difference between DDS and FDS. It is worth highlighting that the Rotations together with the DER group are responsible for more than half of the difference between the DDS and FDS for the gymnasts in group 1 and group 2 of the ranking. (76% and 61% respectively). The Dance Steps is the element that clearly shows smaller differences between the DDS and FDS for the gymnasts placed in the 3 groups of the ranking.

2.4.5. Discussion

According to the results we can see that there are big differences between the scores proposed by the coaches in the competition forms and the difficulties that judges could identify, which increase as the gymnasts go lower in the ranking, going from almost more than 1 point in the gymnasts placed in group 1 of the ranking to almost 4 points for the gymnasts placed in group 3 of the ranking, with a statistically significant difference between the 3 level of gymnasts. Analysing these results we can remark that if we could expect differences between the DDS and FDS in the weaker gymnasts due to

the faults in execution which cancel the value of the difficulty (FIG, 2012), the same result would not be expectable in the very good gymnasts (the first 24 in the ranking), since these gymnasts get higher scores in execution (Levioti, 2012).

The average of the DDS presented in the competition was 9.3 points, close to the maximum grade of 10 points. This average suggests a high global stage of world excellence, but truthfully, as we can see in table 1, not even the best-ranked gymnasts in group one of the ranking can reach this score. Therefore, we can state that the difficulty value of the proposed routine by the coach in the competition form (DDS) is ambitious and does not reflect the performance capacity of the gymnast.

These results could be seen in two ways: on the one hand, perhaps all coaches "overwrite" the competition forms knowing that the judges "feel" the need to cut some difficulties (Ávila, 2011), or, on the other hand, we can also speculate that the lower we go in the gymnasts ranking, the more difficult it is for the judge to identify the difficulty elements, probably due to an execution problem. We propose this point of view because the other possibility is to consider that the judges were not able to identify the difficulties performed by the gymnasts. But, as we know, in Rhythmic Gymnastics World Championship competitions only highly prepared judges can evaluate. Studies such as (St. Marie et al., 2001; Plessener et al, 2005; Johansson, 2010; Dallas, 2010; Heinen, 2012; Fernandez-Villarino, 2013) and more recently (Flessas et al., 2015) have proven that the experience of the judge and her capacity to use other cognitive strategies in perceiving error may be an asset in gymnastics judging. Thus, the non-recognition of the difficulties seems to be inappropriate.

To help explain these results, we tried to understand if these differences between the DDS and FDS could be identified in the different types of difficulty elements. We found out that they exist in all difficulty elements, being significantly higher in the difficulty elements Rotations, DER and Mastery and almost residual in the difficulty elements Dance Steps for the gymnasts in the three groups of the ranking. Analysing the elements first mentioned, these results may suggest the coaches and the judges understand the CoP rules differently. In the difficulty elements with higher differences between the DDS and FDS they may have some problems in the definition of the criteria that characterises them and/or in the comprehension of the technical faults, which cancel the value of the difficulty. On the other hand, and following the same perspective,

in the difficulty elements such as Dance Steps, in which there is a strong proximity to the DDS and FDS, there seems to be an almost perfect understanding of the criteria defined by the CoP.

The contribution of each type of difficulty element to the difference between DDS and FDS is higher in Rotations, DER and Mastery in the three groups of the ranking. These difficulty elements are very complex elements to perform and demand an extraordinary coordination, perfect control of the apparatus technic and a lot of practice hours (Lebre, 2011; Vitrichenko et al, 2011). Therefore, they are also the ones where the gymnast can make more technical faults which cancel the value of the difficulty, mainly the weaker gymnasts. The gymnasts with the intention of getting top scores should present routines with a high level of difficulty combined with good execution quality (Agopyan, 2014). In the case of high ranked gymnasts, this result cannot be explained by the execution scores received, because they were very high. In the case of the gymnasts ranked in groups 2 and 3 (middle and lower gymnasts), the inferior quality in execution may justify these results, suggesting, therefore, that the coaches, do not have a real perception of the performance capacity of their gymnasts in these types of difficulty elements.

In conclusion, the results show there are significant differences between the scores proposed by the coaches and the difficulties that the judges could identify, which increase as the gymnasts go lower in the ranking. The contribution of the difficulty elements to these differences is higher in Rotations, DER and Mastery in the gymnasts placed in the three groups of the ranking. These results, can suggest that the coaches do not have a real perception of the performance capacity of their gymnasts, in these types of difficulty elements.

We suggest that future studies should examine if the difference between DDS and FDS depends on the type of apparatus.

We believe that the evaluation of the difference between DDS and FDS are variables to consider in order to help reconstruct the CoP definitions related to some types of difficulty: (i) the way coaches and judges understand the rules, (ii) the orientation of the training process to maximize the performance capacity of the gymnasts in the type of difficulty elements in which the difference of DDS and FDS is higher, and/or (iii) strategically give preference to the types of difficulties in which this difference is lower.

2.4.6. References

Agopyan A (2014) Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. Journal of Scientific Research 19(12): 1554-1565.

Arkaev LI, Suchilin NG (2004) How to create champions - the theory and methodology of training top-class gymnasts. Oxford: Meyer&Meyer Sport.

Ávila-Carvalho L, Leandro C, Lebre E (2011) 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), Book of abstracts of the 16th Annual Congress of the European College of Sport Science. Liverpool, UK 579-580.

Ávila-Carvalho L, Klentrou P, Palomero ML, Lebre E (2012) Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. The Open Sports Sciences Journal 5: 146-153.

Bucar PM, Cuk I, Pajek J, Kovac M, Leskosek B (2013) Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? Journal of Human Kinetics 37: 173-181.

Caburrasi EF, Santana MV (2003) Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002. Granada2003 Available from: http://www.efdeportes.com/efd65/grd.htm.

Čuk I, Fink H, Leskošek B (2012) Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. Science Gymnastics Journal 4: 73 - 82.

Dallas G, Kirialanis P (2010) Judges' evaluation of routines in men's artistic gymnastics. Science Gymnastics Journal 2: 49-58

Ferreirinha J, Carvalho J, Côrte-Real C, Silva A (2011) Evolução do Valor Real de Dificuldade dos Exercícios de Paralelas Assimétricas de Ginástas de Elite nos Ultimos Ciclos Olímpicos. In FGP (Ed.), Da Prática à Ciência. Artigos do 2º e 3º Congresso de FGP, Lisboa 71-78.

Fernandez-Villarino M, Bobo-Arce M, Sierra-Palmeiro E (2013) Practical Skills of Rhythmic Gymnastics Judges. Journal of Human Kinetics 39: 243-249.

FIG-International Gymnastics Federation (2012) Code of Points for Rhythmic Gymnastics Competitions. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

FIG-International Gymnastics Federation (2013) Gymnastics Results. Available at: http/www.gymnasticsresults.com

Flessas K, Mylonas G, et al. (2015) Judging the Judges' Performance in Rhythmic Gymnastics. Medicine & Science in Sports & Exercise**.** American College of Sports Medicine.

Johansson A (2010) Deciding who is the best. Validity issues in selections and judgements in elite sport. Academic dissertation, Faculty of Social Sciences, Umeå University, Sweden**.**

Heinen T, Vinken P, Velentzas K (2012) Judging Performance In Gymnastics: A Matter Of Motor Or Visual Experience? Science of Gymnastics Journal 4: 63-72.

Hökelmann A, Breitkreutz T, Liviotti G (2012) Changes in performance structure during group competitions in rhythmic gymnastics. Edited by Prof. Derek M. Peters & Dr. Peter G. O'Donoghue, Book of abstracts of the World Congress of Performance Analysis of Sport IX University of Worcester, UK 99.

Lebre E (2011) Technical principles for the new framework. Crossroads to the Future. International Federation of Gymnastics Scientific [Press release].

Liviotti G, Hökelmann A (2012) Which Quantifiable Performance Parameter(s) Determined the Medals Winners at the World Championship 2011 In Rhythmic Gymnastics – Group Competition?. In Rsupesy &T.(Ed.), Book of abstracts of the VI International Scientific Conference of Students and Young Scientists, Modern University Sport Science, Moscow 43.

Massidda M, Calò MC (2012) Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, Journal of Sports Sciences (30)13: 1415-1420.

Pelin RA (2013) Studies Regarding The Rhythmic Gymnastics From The Olympic Games. Sport & Society / Sport si Societate (13): 61-61.

Plessner H (2005) Positive and negative effects of prior knowledge on referee decisions in sports. T. Betsch, Haberstroh,S (Ed.) The routines of decision making 311-324.

Ste-Marie DM, Valiquette SM, Taylor G (2001) Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. Research Quarterly for Exercise and Sport 72(4): 420-426.

Trifunov T, Slobodanka D (2013) The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. Physical Culture 67(2): 120-129.

Vitrichenko N, Klentrou N, Gorbulina N, Della-Chiaie D, Fink H (2011) Groups. In FIG (Ed.), Rhythmic Gymnastics. Technical Manual. Level 3, Lousanne: FIG Academy 3-55.

Wang M, Lu M, Sun X (2013) Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. Journal of Physical Education / Tiyu Xuekan (20) 5: 117.

3. Empirical Studies: The content in Rhythmic Gymnastics high level competition routines

3. Empirical Studies: The content in Rhythmic Gymnastics high level competition routines

3.1. Study V: Technical Content of Elite Rhythmic Gymnastics

Technical Content of Elite Rhythmic Gymnastics

Leandro, C. [1,3], Ávila-Carvalho, L. [2], Sierra-Palmeiro, E. [3], Bobo-Arce, M. [3]

Science of Gymnastics Journal 2015; 8(1), 85-96

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

3.1.1. Abstract

The primary aim of this study was to analyse all technical elements used in the Rhythmic Gymnastics Kiev World Championship 2013, and identify the Structural characteristics of the technical content of elite Rhythmic Gymnastics individual routines. The data has been collected from the difficulty forms concerning 288 individual routines. To allow the comparison between gymnasts with different levels the individual routines were clustered into 3 subgroups according to their final ranking competition. Body difficulty elements were organized, according to the composition requirements stated in the RG Code of Points (FIG, 2012). Non-parametric tests - Kruskal-Wallis, Mann-Whitney and Friedman test were applied to determine whether there were significant differences between groups. As main results we can point out that in general the rhythmic gymnasts used similar body difficulties with limited variety. The highest valued elements are Dynamic Elements with Rotation and throw (DER) and rotations and these represent 50% of the total value of the routine. Concerning the dance steps and mastery, no differences were found between the routines of gymnasts place in the three parts of the ranking. The routines had differences in the composition pattern between the gymnasts according to the final ranking of the gymnasts in following items: (i) on the number of rotations of flat foot or other part of the body, Fouetté rotations and Mixed Difficulties; (ii) on the value of jumps, rotations, DER and Mixed Difficulties.

Key Words: Body difficulty, individual routines, evaluation, rhythmic gymnastics.

3.1.2 Introdution

The main reason for the success in RG competition is the capacity to perform the exercise, with high level body elements and apparatus technic, with perfect execution, in harmony with the character and rhythm of the music, respecting the principle of originality and diversity. This is a guarantee of an exciting high performance sport to watch.

The rules which guide the routines composition can also have influence in the gymnasts' performance (Massidda, 2012).

The limited variety on the choice in difficulty elements makes the routine composition boring and puts in risk its artistic value (Ávila, 2012a).

The skilful interaction between the gymnast and the apparatus and the increase difficulty elements in the routines composition are the development in RG (Lebre, 2011).

The analysis of these factors can, according to Ávila, (2012b), influence the developmental programs for the practice and the experimental designs used in the scientific research in RG. The knowledge can also contribute to preview and characterize the effort requirements allowing improvements in the gymnasts' preparation to the competition readiness (Ferreirinha, 2009).

Ferreirinha (2009) refers that to determine the training models it is important to know the characteristics competition routines for high level gymnasts including the details concerning the specificity of their components.

Is, than, fundamental to analyse the development tendencies for the sport in general and to identify specificities of each component as we propose to do with the structural characteristics of the difficulty elements including the diversity and variety in the routines.

The routines composition is not stable concerning their content because they have to be adapted to changes done in the Code of Points (CoP) every Olympic cycle. An individual RG routine is composed by a series of body and apparatus elements linked in a specific way which we call difficulty elements (D). On the present Olympic cycle, the content of and individual RG routine should respect the specific requirements that are common to the routines of all 4 apparatus: jumps/leaps, balances, rotations, Apparatus Mastery, Dance Steps and Dynamic Elements with Rotation and throw (DER) (FIG, 2012). The value of each difficulty element is from 0.10 points to 1.50 points or more, which

may be absolutely determinant in the final score obtained in competition. The inclusion of complex abilities in the routines is essential to have a high score (Massida, 2012).

The CoP holds a great variety of difficulty elements to be used in the routines. One important characteristic of the RG is to allow the gymnast to link it in her own way, with a stylish presentation, clever configuration, and prefect presentation (Wang, 2013). An eventual lack of variety in the body difficulty included in the routines can cause judges and audience dissatisfaction from the point of originality and variety. RG is a visually appealing sport, thus, it is very important to keep the high interest of the public (Agopyan, 2014).

The studies published concerning the content of the RG routines (Caburrasi, 2003; Bobo, 2010; Ávila, 2011; Ávila, 2012; Trifunov, 2013; Agopyan, 2014), include the analysis of the number and the level of difficulty elements but they have no information about contribution of each type of element for the final D score. Also in consequence of the constant and quickly evolution of this sport, a permanent upgrade of these kind of studies are needed (Caburrasi, 2003; Cuk, 2012; Massidda, 2012; Hökelmann et al., 2012; Bucar, 2013, Pelin, 2013).

Within this context, the main goal of the present study is to identify the difficulty elements included in the routines that contribute the most to the success in competition and to analyse the diversity of the body movements included in the difficulty elements. The present study can have an important contribution for the coaches mainly to: supporting the coaching process, defining performance profiles for individual gymnasts, ranking performances, creating data bases in order to identify the most influencing performance indicators and the tendencies in the development of RG (Liviotti, 2012).

3.1.3. Methods

Subjects and design

288 difficulty forms concerning individual routines were analysed. The routines were performed by gymnasts from 45 different countries competing at Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013.This study was done with the permission of the International Gymnastics Federation (FIG).

The official Difficulty forms, submitted prior to the competition, included the routine compositions recorded using the RG CoP symbols. All difficulty elements

reported in the difficulty forms were analysed. The analyse was done considering the all sample, and the sample clustered into 3 subgroups according to gymnasts final ranking as follows: the first part of the ranking - the top 24 gymnasts, the second part of the ranking - 24 middle gymnasts and third part of the ranking – the 24 lower placed gymnasts on the ranking, to allow the comparison the technical elements within gymnasts of different levels.

The analysis was conducted by two international RG judges. The intraclass correlation coefficient (ICC) in test-retest method (intra-examiner) was 0.99. The ICC between the observers (inter-examiner) was 0.98.

Statistical analysis

The data were analyzed using the Statistical Package for Social Sciences – version 20.0 (SPSS 20.0, Chicago, USA) and Microsoft Office Excel 2007. The level of significance was set at $\alpha = 0.05$ (confidence interval of 95%). Descriptive statistics were calculated using the mean values as a measure of central tendency, standard deviation (SD) as a measure of dispersion, and minimum and maximum as measures of data range. After checking the abnormalities in the data distribution ($p<0.05$) using the Kolmogorov-Smirnov normality test, we resorted to non-parametric test (Kruskal-Wallis, Mann-Whitney and Friedman test) to determine whether there were significant differences between the three subgroups in the Rhythmic Gymnastics World Championship ranking. A multiple regression was used to analyze the influence of each difficulty element in the gymnasts' final difficulty score.

3.1.4. Results

The difficulty elements reported in the individual routines were grouped by technical categories: balances, jumps, rotations, masteries, dance steps, and DER, mixed difficulties (MixDif) and criteria associated to difficulty (waves and pre-acrobatics). The results for each category are presented both quantitatively (number of occurrences) and qualitatively (technical value and type) in Figure 1. From Figure 1 we can highlight the number of the mastery (4.0±2.80) and the value of the rotations with 29% of the total value of the routine (2.7±0.83 points). When we observe the three difficulty groups that are based on the body movements (jumps, balances and rotations) we can see that the

rotations have the higher number (3.3±0.61) and the balances the lower number (2.4±1.00). Concerning the rotations, the gymnasts included preferably those with 0.30 points values in their routines. Between them it is possible to highlight the “pivot attitude” (0.52±0.50), the “pivot free leg in ring in back with help” (0.42±0.50) and the “rotation penché” (0.76±0.43). The most used jumps were those with 0.5 points value, mainly the “jeté with turn” (0.82±0.80) and the "jeté with a turn with back bend" (0.45±0.53); The balances with base value 0.5 points were the most performed by the gymnasts, mainly the balance “side scale with split, without help” (0.44±0.49) and balance "back scale leg high up" (0.40±0.49). The most used MixDif were the link of the balance “front scale with back split” and “ring without help” (0.15±0.52). For DER, the most used criteria to raise the value were: “change of level”, “change of body rotation axis”, “throw/catch outside of visual control” and “throw/catch without the help of the hands”.

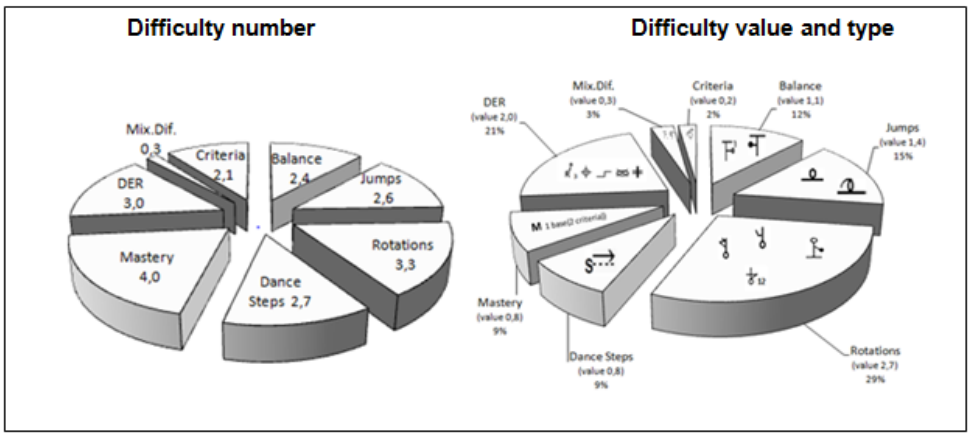


Figure 1. Difficulty elements presented in the Rhythmic Gymnastics individual routines in the 2013 World Championships clustered according to number, value and type.

Analysing the sample according to final ranking of the gymnasts, significant differences were found on the number of balances, MixDif, rotations on the flat foot or other part of the body, and “fouetté” rotations (Table 1). No other significant differences in the number of technical difficulties were found according to the final ranking of the gymnasts (Figure 2).

The number of balances was significantly higher in the gymnasts of the 3rd part of the ranking and the MixDif significantly higher in the gymnast of the 1st part of the ranking. The number of rotations on the flat foot or other part of the body is higher in

the 1st part of the ranking and decreases significantly in the 2nd and 3rd parts. The number of "fouetté" rotations is significantly higher in the gymnast of the 2nd part of the ranking (Table 1).

Table 1. Number of balances, MixDif and Rotations in the Rhythmic Gymnastics individual routines clustered according to gymnasts' final ranking in the 2013 World Championships.

	1st part of the ranking (n=96)		2nd part of the ranking (n=96)		3rd part of the ranking (n=96)		Kruskal-wallis test	Pairwise Comparisons
Difficulty number	**Mean+sd**	**Min-Max**	**Mean+sd**	**Min-Max**	**Mean+sd**	**Min-Max**	**P=**	
Balance	2,16±1,08	0-4	2,48±0,91	1-4	2,66±0,93	0-4	0,002*	rk1-rk3 0,001
Mix. Dif.	0,75±0,97	0-2	0,56±0,90	0-2	0,38±0,78	0-2	0,016*	rk3-rk1 0,012
Rot.flat foot..	1,03±0,49	0-2	1,00±0,50	0-2	0,68±0,53	0-2	0,000*	rk3-rk1 0,000 rk3-rk2 0,000
Rot. "Fouette"	0,28±0,49	0-2	0,65±0,69	0-2	0,58±0,57	0-2	0,000*	rk1-rk3 0,001 rk1-rk2 0,000

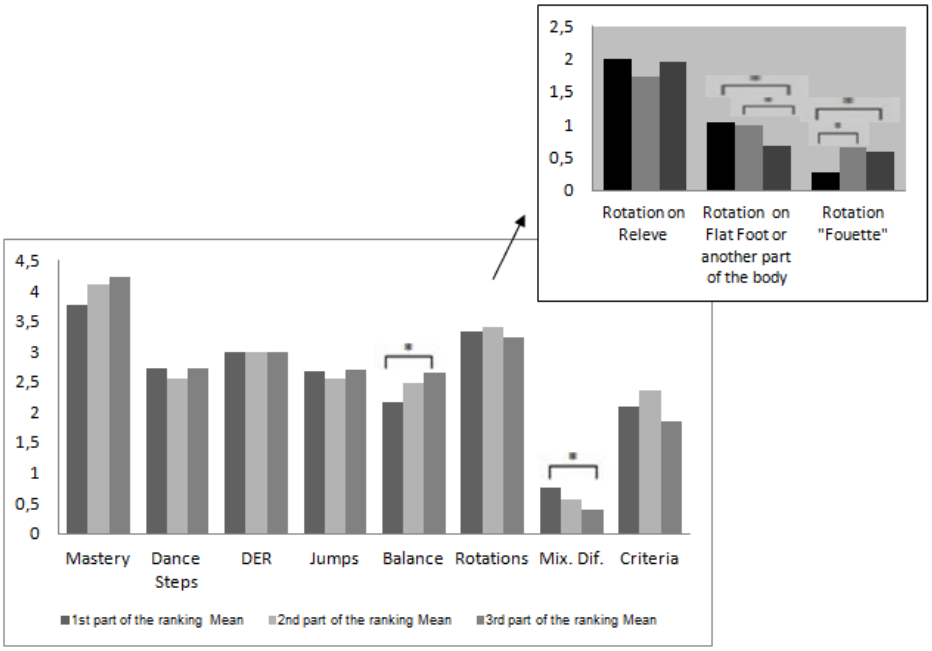


Figure 2. Average number of difficulty elements presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking.(*p<0.05)

The value of the DER, rotations, jumps and MixDif is higher in the gymnast placed in the 1st part of the ranking and decreases significantly in the 2nd and 3rd parts. For mastery, dance steps, balances and criteria associated to difficulty (waves and pre-acrobatics) there were no statistically significant differences regarding the technical

value, and the gymnasts' final ranking (Table 2). Concerning the value of the rotations we can highlight that the fouetté rotations had a significant higher value in the gymnasts placed on the 2nd part of the ranking when compared to the gymnast in the 1st part. The rotations of flat foot or another part of the body registered a higher value in the 1st part of the ranking (Figure 3).

Table 2. Technical value of DER, Jumps, MixDif and Rotations presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking.

	1st part of the ranking (n=96)		2nd part of the ranking (n=96)		3rd part of the ranking (n=96)		Kruskal-wallis test	Pairwise Comparisons
Difficulty value	Mean+sd	Min-Max	Mean+sd	Min-Max	Mean+sd	Min-Max	P=	
DER	2,05±0,30	1,3-2,7	2,01±0,27	1,4-2,8	1,83±0,35	1,0-2,8	0,000*	rk3-rk2 0,001 rk3-rk1 0,000
Jumps	1,54±0,44	0,7-2,8	1,4±0,45	0,7-2,7	1,25±0,33	0,5-2,3	0,000*	rk3-rk1 0,000
Mix. Dif.	0,37±0,49	0,0-1,7	0,25±0,41	0,0-1,0	0,16±0,34	0,0-1,0	0,005*	rk3-rk1 0,003
Rotations	3,00±0,88	1,2-4,7	2,89±0,66	1,2-4,3	2,22±0,71	0,7-3,9	0,000*	rk3-rk2 0,000 rk3-rk1 0,000
Rot. Releve	1,85±0,97	0,0-4,0	1,56±0,88	0,0-3,3	1,37±0,67	0,0-3,1	0,001*	rk3-rk1 0,001
Rot.flat foot..	0,89±0,44	0,0-2,3	0,80±0,37	0,0-1,7	0,49±0,40	0,0-1,4	0,000*	rk3-rk2 0,000 rk3-rk1 0,000
Rot.Fouette	0,25±0,47	0,0-2,0	0,53±0,59	0,0-2,0	0,36±0,38	0,0-1,2	0,001*	rk1-rk2 0,001

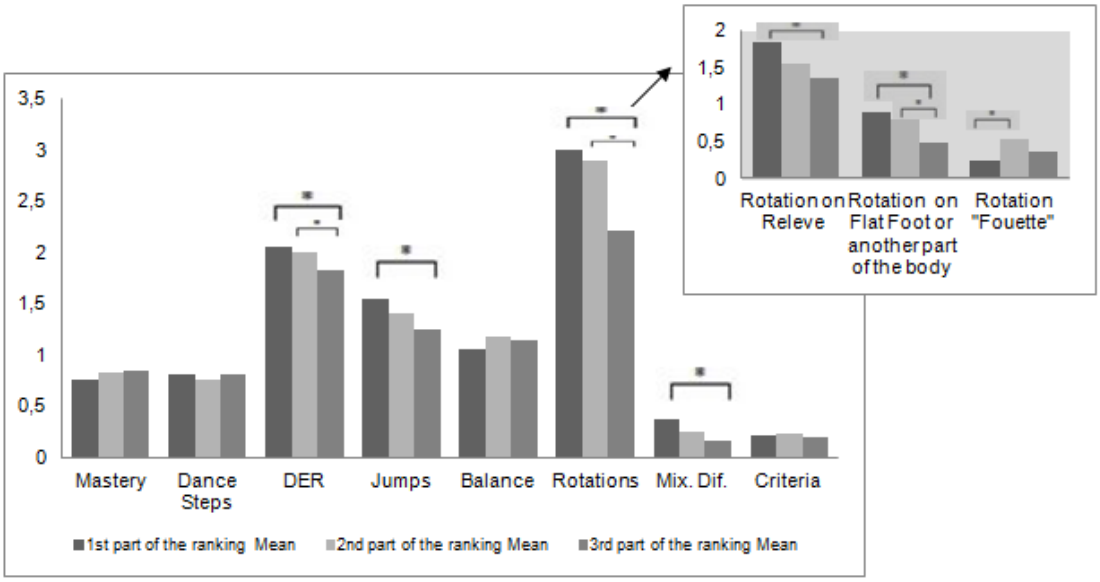


Figure 3. Technical value of the difficulty elements presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking. (*p<0.05)

Jumps

Analysing the value of the jumps included in the routines according to the final ranking of the gymnasts, we found significant differences for all jumps except the jumps with 0.5 points value. The routines of the gymnasts placed in 1st part of the ranking had a higher number of jumps 0.7 and 0.8 points value. The jumps of value 0.2, 0.3 and 0.4 points were the preferred of gymnasts placed in the 3rd part of the ranking. The jumps with 0.6 points value are performed preferably by the gymnasts on the 2nd part of the ranking. The jumps with 0.5 points value jumps were the preferred of all gymnasts independently of their place on the final ranking. There were not significant differences for the gymnasts ranking regarding the jumps of 0.5 value jumps (Figure 4).

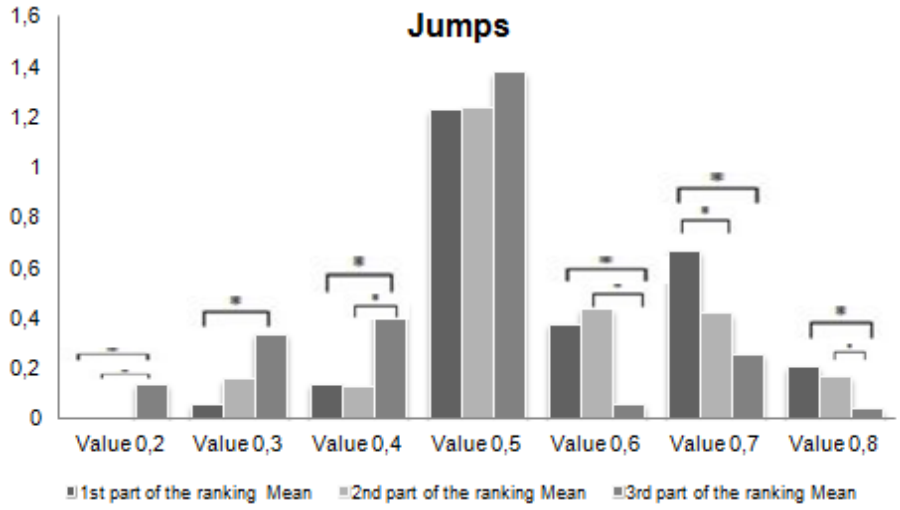


Figure 4. Number of Jump difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships.

Kruskal-wallis test * p<0,05

Balances

We found significant differences in the number of balance difficulties when we compare the routines performed by the gymnasts of different parts of the final ranking. The gymnasts ranked in the 3rd part of the ranking had a higher number of balances with 0.30 and 0.40 points value in their routines. On the other hand the gymnasts ranked on the 1st and 2nd parts preferred to include in their routines balances of 0.50 points (figure 5).

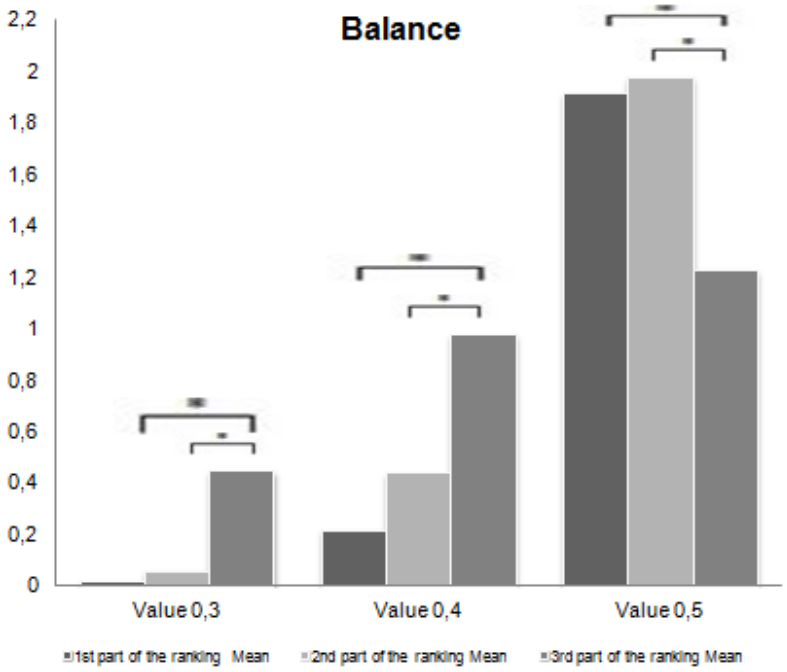


Figure 5. Number of balance difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships.

Kruskal-wallis test * p<0,05

Rotations

There were significant differences in the rotations included in the routines in all parts of the final ranking excepted for the rotation of 0.5 points value. The routines of the gymnasts ranked in the 3rd part had a higher number of rotations on "relevé" of 0.1, 0.2 and 0.4 points value. On opposite, the gymnasts placed in the 1st and 2nd parts preferred to include 0.5 and 0.6 points value rotations on "relevé" in their routines. The rotations on relevé with 0.3 points value were the most performed by all gymnasts independently of their position in the final ranking (Figure 6). Concerning the rotations on flat foot or another part of the body, it was clear that the gymnasts placed in the 1st and 2nd parts of the ranking preferred to include this type of rotation with 0.4 points value in their routines.

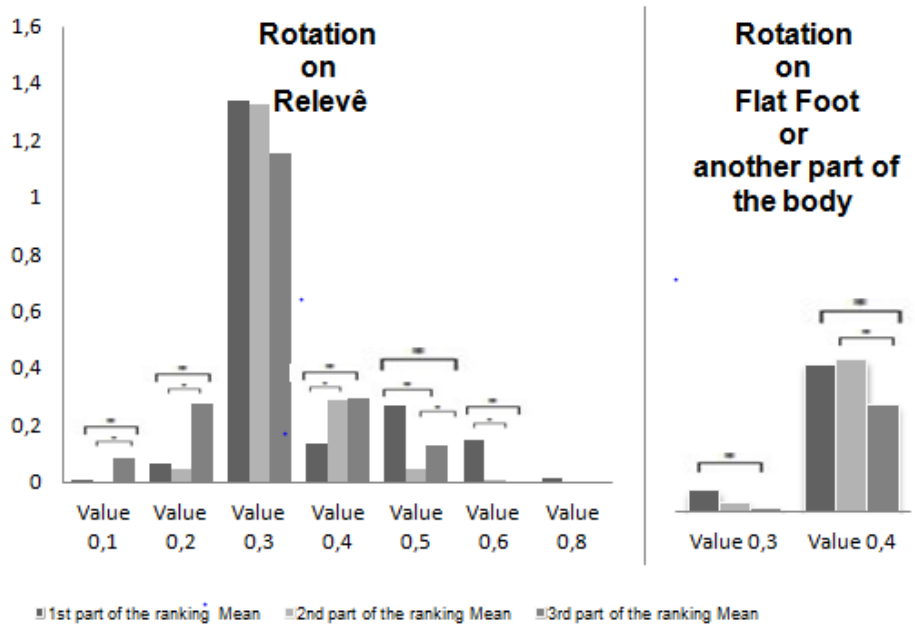


Figure 6. Number of Rotations difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships.

3.1.5. Discussion

This study provides a quantitative and qualitative analysis of the difficulty elements used in the individual routines of the 2013 RG World Championships.

The 288 individual routines studied were clustered into three subgroups according to the gymnasts' final ranking in the World Championships. We discussed the results (number, value and type) in 3 dimensions: (1) global analysis of the composition of the routines; (2) analysis by group of difficulty elements; (3) ranking of the gymnasts.

In a global point of view the routines hold an average value of 9.30 points, very close to the maximum possible score of 10 points. Despite the World Championships being one of the most important competition in the calendar, this result may lead to a false analysis, as it could mean such a high a plateau of international excellence which in reality only occurs amongst gymnasts at the top of the ranking. Ávila, (2011) studied the difference between the departure score (presented by the coach in the difficulty form) and the final score obtained by the gymnast and concluded that the majority of the gymnasts reach very significant differences, of 2 or more points between these two scores.

We also highlight the fact that the rotations and the DER, together represented 50% of the total value of the composition. This result showed an important change in the global content of the routines in this Olympic Cycle. Studies such as Caburrasi (2003) and Ávila (2011) showed that in the previous Olympic cycles the highest contribution in the routines value came from the Jumps. The increase in the rotations and DER values happens because in the present Olympic cycle it is possible to add some criteria to these difficulties that allow the gymnast to increase its value and degree of complexity (FIG, 2012). These results can be analysed in two different perspectives. On the one hand, it represents an upgrade of the execution quality, but on the other hand, it means that an extreme importance is given to 2 types of difficulty elements leading to an under estimation of the other groups. We also remarked the lack of variety and diversity in the elements chosen that has been repeatedly mentioned in previous studies concerning individual routines (Bobo, 2003; Agopya, 2014) and group routines (Ávila, 2011b; Ávila 2012; Ávila, 2012b). Therefore it is possible to conclude that the RG routines present a consistent pattern in the usage of the difficulty elements.

The type of difficulty elements used in the routines is similar, with some difficulty elements being repeated several times in the routines. This means that the routines' composition is not defined by being unique, with diversity and creativity, characteristics that are necessary for the enrichment of the routines composition (Balcells, 2009; Leandro, 2015) and reflect the spectacular of the choreography (Pelin, 2013).

The analysis according to the type of difficulty showed us similar results as Agopyan (2014) for routines performed on the last Olympic cycle: the rotation difficulties (mainly the "relevé" rotations) were the preferred of the gymnasts and the balance difficulties the less used. The rotation difficulties are very complex elements to perform (Lebre, 2011; Vitrichenko et al, 2011), but they are also those where the gymnast can get more points, once the CoP (FIG, 2012) allows to add the base value of the difficulty for each rotation performed. The lower number of balances in the routines is, probably, due to the fact that the gymnasts spend considerable time of the routine to perform them because they are static difficulty elements (Gateva, 2015) and they have low values: 0.50 points is the maximum possible value for a balance, according to the CoP (FIG, 2012). These are the main reasons for the preference of the gymnasts to include more difficulties in rotation and less in balance in their routines. The routines

only last for maximum 90 seconds and they have to optimize the time available to get the maximum of points allowed (10 points). The gymnasts, with the intention of getting top scores should present routines with a high level of difficulty combined with good execution quality (Agopyan, 2014).

The mastery and dance steps have comparatively lower possible values than the jumps, rotations and balances. These groups have an inferior degree of execution complexity, they are less valued in the CoP (FIG, 2015). To promote the inclusion of these types of elements in RG routines, and therefore have more interesting choreographies their value should be increased (Livotti, 2012; Leandro, 2015). One of basic requirements of RG is that the gymnast should show an optimal use of the body together with the apparatus handling. In this way, to raise the difficulty departure score the gymnast must increase of both body and apparatus difficulty level included in the routine (Agopyan, 2014).

The analysis of the results according to the gymnasts' final ranking showed that the higher placed gymnasts chose preferentially elements with a higher complexity (MixDif, rotations on flat foot or other part of the body and "fouetté" rotations) and the lower placed gymnasts chose elements with lower complexity (balances) as described also by Gateva (2015).

Regarding the difficulty elements value, the jumps were the elements with higher value for the gymnasts in the first and second parts of the ranking. With the exception of the jumps of 0.5 points value, the gymnasts higher placed (1st and 2nd part of the ranking) include preferably the jumps with higher value and the gymnast placed on 3rd part preferred the jumps of 0.3 and 0.4 points value, which confirms the expectable. According to Bobo (1998) and Bobo, (2003), as a norm the best gymnasts hold physical and artistic capacities that allow them to perform more and higher level elements with high execution complexity. The rotations, DER and MixDif had higher values in the routines of the gymnasts placed in 1st part of the ranking and decreased in the routines of the gymnasts placed in the second and third parts. The complexity of this type of difficulties is very high and demands an extraordinary coordination, a perfect control of the apparatus technic and a lot of practice hours, (Lebre, 2011; Vitrichenko et al, 2011), which justifies that they are preferably used by the gymnasts highly ranking.

3.1.6. Conclusions

The rhythmic gymnasts who competed at the 2013 World Championships used in their routines very similar difficulties elements with limited variety. The more used difficulties were the rotation “attitude”, rotation with “free leg in ring in back with help”, “rotation in penché”; balance “side scale with split, without help” and balance “back scale leg high up”; jump “jeté with turn” and “jeté with a turn with back bend”.

The highest valued elements are DER and rotations and these represent 50% of the total value of the routine. These groups showed an important contribution to the final D score. The balances were the less used difficulty group.

The routines had differences in the composition pattern between the gymnasts according to the their final ranking in the following items: (i) the number of rotations of flat foot or other part of the body, “fouetté” rotations and MixDif; (ii) the value of jumps, rotations, DER and MixDif. Concerning the dance steps and mastery, no differences were found between the routines of gymnasts place in the three parts of the ranking.

This study provides updated information about the individual routines content in rhythmic gymnastics, to be considered: (i) to the possible modifications of the present Code of Points, in particular for the definition of the composition requirements in order to have higher level of variety and diversity in the routines, and (ii) to the training process to achieve the high performance level in the individual gymnasts.

3.1.7. References

Agopyan, A. (2014). Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. Journal of Scientific Research, 19(12), 1554-1565.

Arkaev, L.I.&Suchilin, N.G.(2004). How to create champions - the theory and methodology of training top-class gymnasts. Oxford: Meyer&Meyer Sport.

Ávila-Carvalho, L., Klentrou, P., Palomero, M. d. L., & Lebre, E. (2012) Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. The Open Sports Sciences Journal, 5, 146-153.

Ávila-Carvalho, L., Klentrou, P., & Lebre, E. (2012). Handling, Throws, Catches and Collaborations in Elite Group Rhythmic Gymnastics. Science of Gymnastics Journal, 4 (3), 37-47.

Ávila-Carvalho, L., Leandro, C., & Lebre, E. (2011). 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), Book of abstracts of the 16th Annual Congress of the European College of Sport Science. Liverpool, UK. 579-580.

Ávila-Carvalho, L., Palomero, M. d. L., & Lebre, E. (2011b). Estudio del valor artístico de los ejercicios de conjunto de Gimnasia Rítmica de la Copa del Mundo de Portimão 2007 y 2008. Apunts. Educación Física y Deportes, 1.er trimestre(103), 68-75.

Ávila-Carvalho, L., Palomero, M. d. L., & Lebre, E. (2009c). Difficulty score in Group Rhythmic Gymnastics. Portimão 2007/2008 World Cup Series. Palestrica Mileniului III. Civilizatie si sport, Anul X, 3(37), 261-267.

Balcells, M., Martín, C., & Anguera, M. (2009). Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión Corporal y Danza Contact-Improvisatio. Apunts educación física y deportes.Ciencias aplicadas a la actividad física y el deporte, 14-23.

Bobo M, Sierra E.(2003). Estudio de las repercusiones de los cambios de código de puntuación en la composición de los ejercicios de gimnasia rítmica en la técnica corporal. Available from: http://www.cienciadeporte.com/congreso/04%20val/pdf/p3.pdf.

Bobo M, Sierra E. (1998). Una nueva propuesta de dificultades corporales en gimnasia rítmica deportiva. Libro de resúmenes del VI Congreso de Educación Física e Ciencias do Deporte dos Países de Lingua Portuguesa Deporte e Humanismo en clave de Futuro.

Bucar, P. M., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? Journal of Human Kinetics, 37, 173-181.

Caburrasi, E.F., Santana, M.V. (2003). Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002. Granada2003 Available from: http://www.efdeportes.com/efd65/grd.htm.

Čuk, I., Fink, H., & Leskošek, B. (2012). Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. Science of Gymnastics Journal, 4, 73 – 82.

FIG. (2012). Code of Points for Rhythmic Gymnastics Competitions. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

Ferreirinha,J., Carvalho,J., Côrte-Real, C., Silva, A. (2011).Evolução do Valor Real de Dificuldade dos Exercícios de Paralelas Assimétricas de Ginástas de Elite nos Ultimos Ciclos Olímpicos. In FGP (Ed.), Da Prática à Ciência. Artigos do 2º e 3º Congresso de FGP (pp. 71-78). Lisboa.

Gateva, M., Gospodarski, N., Treneva, V., Avramov, D., Ivanov, N., Andonov, K. (2015). Comparison Between The Static Balance Of Practitioners From Different Sports and Non-Athletes. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. Book of abstracts of the 20th Annual Congress of the European College of Sport Science. p. 569. Malmo, Sweden.

Hökelmann, A., Breitkreutz, T., & Liviotti, G. (2012). Changes in performance structure during group competitions in rhythmic gymnastics. Edited by Prof. Derek M. Peters & Dr. Peter G. O'Donoghue, Book of abstracts of the World Congress of Performance Analysis of Sport IX University of Worcester, UK. p99.

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines. Science of Gymnastics Journal. (Accepted for publication 12 July 2015).

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). What Do Rhythmic Gymnastics Judges Think About Their Code Of Points?. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. Book of abstracts of the 20th Annual Congress of the European College of Sport Science (pp. 569). Malmo, Sweden.

Lebre, E. (2011). Technical principles for the new framework. Crossroads to the Future. International Federation of Gymnastics Scientific [Press release].

Liviotti, G. Hökelmann, A. (2012). Which Quantifiable Performance Parameter(s) Determined the Medals Winners at the World Championship 2011 In Rhythmic Gymnastics – Group Competition?. In Rsupesy &T.(Ed.), Book of abstracts of the VI International Scientific Conference of Students and Young Scientists "Modern University Sport Science" . Moscow, p43.

Liu, X.X., Kuang, L., (2001) Review of evolvement course of international evaluation rules in rhythmic gymnastics and its effects on technique development. Journal of Beijing Sport University. (24)3. 412-415

Massidda, M. & Calò, M.C. (2012). Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, Journal of Sports Sciences, (30)13, 1415-1420

Pelin, R. A.(2013) Studies Regarding The Rhythmic Gymnastics From The Olympic Games. Sport & Society / Sport si Societate, (13), p61.

Sands, w.A.,Caine, D.J., &Bornes, J. (2003). Scientific aspects of women´s gymnastics (vol.45). Basel: S. Karger A.G.

Trifunov, T., & Slobodanka, D. (2013). The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. Physical Culture, 67(2), 120-129.

Vitrichenko, N., Klentrou, N., Gorbulina, N., Della Chiaie, D. & Fink, H. (2011). Groups. In FIG (Ed.), Rhythmic Gymnastics. Technical Manual. Level 3. 3-55. Lousanne: FIG Academy.

Wang,M., Lu,M., Sun, X. (2013). Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. Journal of Physical Education / Tiyu Xuekan, (20) 5, p117.

3.2. Study VI: Análisis Cualitativo y cuantitativo de los Ejercicios Individuales de Gimnasia Rítmica en diferentes aparatos: Variedad y Diversidad.

*Quantitative and qualitative analyses of the Rhythmic Gymnastics Individual Routines in the different apparatus: Variety and Diversity.*Leandro, C.[1,3], Ávila-Carvalho, L.[2], Sierra-Palmeiro, E.[3], Bobo-Arce, M.[3] Revista apunts Educatió Fisica e Esports 2015; (Accepted for publication 20 april 2016)

[1] *University Lusófona of Porto – Faculty of Psychology, Education and Sport (Portugal)*
[2] *Porto University – Sports Faculty* - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto *(Portugal)*
[3] *University of Coruña – Faculty of sport Science and Physical Education (Spain)*

3.2.1. Resumen

El objetivo de este estudio es hacer un análisis cuantitativo y cualitativo de los elementos técnicos utilizados en ejercicios individuales, en el Campeonato del Mundo de Gimnasia Rítmica (GR) en Kiev en 2013 e identificar las características estructurales de la composición, que permiten percibir la variedad y diversidad en los ejercicios de GR en cada uno de los aparatos. Se analizaron 288 fichas de la competición, correspondientes a los ejercicios de 72 gimnastas en cada aparato (Aro, Pelota, Mazas y Cinta). Los datos fueron analizados mediante estadística descriptiva y pruebas no paramétricas (Kruskal-Wallis, Mann-Whitney y prueba de Friedman). Los resultados muestran características estructurales en la composición de los ejercicios muy similares en los diferentes aparatos respecto a los aspectos de la técnica corporal que determinan el mayor porcentaje en la composición, Los ejercicios de los diferentes aparatos se distinguen en los grupos más relacionados con la parte artística (Maestría, Pasos rítmicos y DER), principalmente debido a las características específicas de cada aparato. Este estudio proporciona información actualizada sobre el contenido técnico de los ejercicios individuales de GR de élite, para ser considerado respecto a: (i) la posibilidad de modificar el presente Código de Puntuación, sobre todo en la definición de las exigencias de composición que favorezcan la variedad y diversidad y fomentar el valor artístico y la unidad técnica del aparato; (ii) el proceso de formación de valor y el perfil de desempeño de GR gimnastas de elite.

Palabras Clave: Técnica corporal, Técnica de aparato, Variedad, Diversidad, Gimnasia Rítmica

3.2.2. Introducción

El principal determinante del éxito en la competición en gimnasia rítmica (GR) es la capacidad de llevar a cabo con corrección máxima, los elementos de elevado nivel de técnica corporal y de aparato, en perfecta armonía con el carácter y el ritmo de la música. Este supuesto debe dar como resultado una coreografía que, por su originalidad y diversidad se presente como un magnífico espectáculo de arte para el espectador y sea visto como un desafío en el alto rendimiento.

La composición de los ejercicios se caracteriza por ser única, con diversidad y creatividad (Balcells, 2009), lo que refleja el espectáculo de coreografía (Pelin, 2013). En este ciclo olímpico, el Código de Puntuación (CP) (FIG, 2012) presenta de forma inteligente y de fácil percepción, una amplia gama de opciones en cuanto a la elección de los diferentes elementos de técnica corporal (Wang, 2013) que pueden formar parte de la composición de los ejercicios de GR por decisión del entrenador (Vitrichenko et al, 2011) y cumplir con los requisitos específicos establecidos para los aparatos (Aro, Pelota, Mazas y Cinta) en gimnastas de alto nivel.

La gimnasia rítmica ha experimentado una evolución constante y espectacular de su técnica en los últimos años debido principalmente a la evolución de su Código de Puntuación (CP) que ha estado buscando una mayor apreciación de los ejercicios de competición (Sierra, 2015).

Es importante examinar si estos requisitos son línea orientadora en la evolución de este deporte, que contemple el principio del espectáculo deportivo, tanto por la variedad y diversidad de elementos técnicos representados en cada ejercicio y para cada aparato, como por el aumento de la complejidad de la aplicación de esos mismos elementos.

La falta de variedad y similitud de los elementos de técnica corporal en diferentes aparatos en gimnastas de diferentes niveles puede causar insatisfacción para el público en el aparatado de la originalidad y la variedad. (Agopyan, 2014)

Considerando que los requisitos de composición establecidos en el CP, tienen una base común en cuanto el número y valor en todos los aparatos con respecto a los elementos de técnica corporal (Equilibrios, Saltos y Giros) y el número y el valor de las Maestrías, Pasos rítmicos y de Elementos de Riesgo (DER), es esencial percibir si estas definiciones nos guían para construir ejercicios de GR demasiado similares en la misma

gimnasta en cada aparato y entre las diferentes gimnastas; si esto compromete la variedad y la diversidad en cada ejercicio en los diferentes aparatos y en última instancia, el espectáculo deportivo. El resultado del análisis de estos factores puede, según Ávila (2012), que influya en los programas de desarrollo de la práctica, así como los diseños experimentales utilizados en la investigación científica en GR.

Este tipo de estudios deberían influir en la construcción del nuevo CP que aparece cada ciclo olímpico. En este contexto, el objetivo de este estudio es hacer un análisis cuantitativo y cualitativo de los elementos técnicos utilizados en ejercicios individuales, en el Campeonato del Mundo de Gimnasia Rítmica en Kiev en 2013 e identificar las características estructurales que permiten percibir la variedad y diversidad en los ejercicios de GR en cada uno de los aparatos.

3.2.3. Material y Métodos

Se analizaron 288 fichas de la competición, utilizadas en el Campeonato Mundial de GR en Kiev en 2013, correspondientes a los ejercicios de 72 gimnastas en cada aparato (Aro, Pelota, Mazas y Cinta).

El análisis de los elementos de dificultad corporal (D) se organizó de acuerdo con los requisitos de composición de CP 2012/2016 sobre Equilibrios, Saltos, Giros, las Dificultades mixtas, Maestría, Pasos Rítmicos y Elementos dinámicos de rotación con lanzamiento (DER).

Este estudio fue aprobado por la Federación Internacional de Gimnasia (FIG).

Se realizó un análisis estadístico con SPSS y Excel programa de 2010. El análisis fue realizado por dos jueces internacionales de gimnasia rítmica. El coeficiente de correlación intraclase (CCI) en el método test-retest (intra-examinador) fue de 0,98.

La CPI entre los observadores (inter-examinador) fue de 0,97. Los datos fueron analizados mediante estadística descriptiva y pruebas no paramétricas: Kruskal Wallis, Mann-whitney y Prueba Friedman.

3.2.4. Resultados

Los elementos de dificultad observados en ejercicios individuales se agrupan de acuerdo con el código de puntuación de 2012 en las siguientes categorías técnicas: Equilibrios, Saltos, Giros, Maestría, Pasos Rítmicos, Elementos dinámicos con rotación y Lanzamiento (DER), Dificultad mixta (DifMix), y Criterios relacionados con la dificultad (ondas y preacrobacias).

Los resultados para cada tipo de elemento, se presentan cualitativamente (valor técnico y el tipo de dificultad) y cuantitativamente (frecuencias). Respecto al valor global (total de elementos de Dificultad) de la composición se obtiene una media y una desviación típica en Aro (9,37 ± 0,72); Pelota (9,30 ± 0,85); Cinta (9,14 ± 0,94) y mazas (9,30 ± 0,79).En cuanto al valor aportado por cada tipo de los elementos de dificultad podemos destacar que el valor más alto corresponde al grupo de los Giros y los DER en todos los aparatos.

Al considerar los tres grupos de dificultad, base del cuerpo (Salto, Equilibrios y Giros) destacan con mayor frecuencia el de los Giros en Aro, Pelota, Cinta y Mazas (2,51 ± 0,76, 2,74 ± 0,88, 2,80 ± 0,85; 2,76 ± 0,81), respectivamente, y el de menor frecuencia el grupo de los Equilibrios en Aro, Pelota, Cinta y Mazas (1,19 ± 0,48, 1,15 ± 0,47, 1,02 ± 0,47, 1,12 ± 0,50), respectivamente. (Figura 1).

Se encontraron diferencias significativas en los diferentes aparatos en el valor de la dificultad de la Maestría, Pasos Rítmicos, DER y criterios asociados a la Dificultad (Tabla 1), especialmente en el valor de los elementos con maestría entre la Cinta y el Aro (p = 0:00), Cinta y Pelota (p = 0:00), Mazas y Aro (p = 0,01) y Mazas y Pelota (p = 0:00); el valor de los Pasos Rítmicos entre Pelota y Mazas (p = 0:00), Pelota y Cinta (p = 0:00), Aro y Mazas (p = 0,01) y Aro y Cinta (p = 0:00); el valor DER entre Pelota y Mazas (p = 0:00), Pelota y Aro (p = 0:00), Cinta y Mazas (p = 0:00) y Cinta y Aro (p = 0:00) y los Criterios de valor asociado con las dificultades entre Aro y Mazas (p = 0,05) y Aro y Pelota (p = 0,01) (Figura 1).

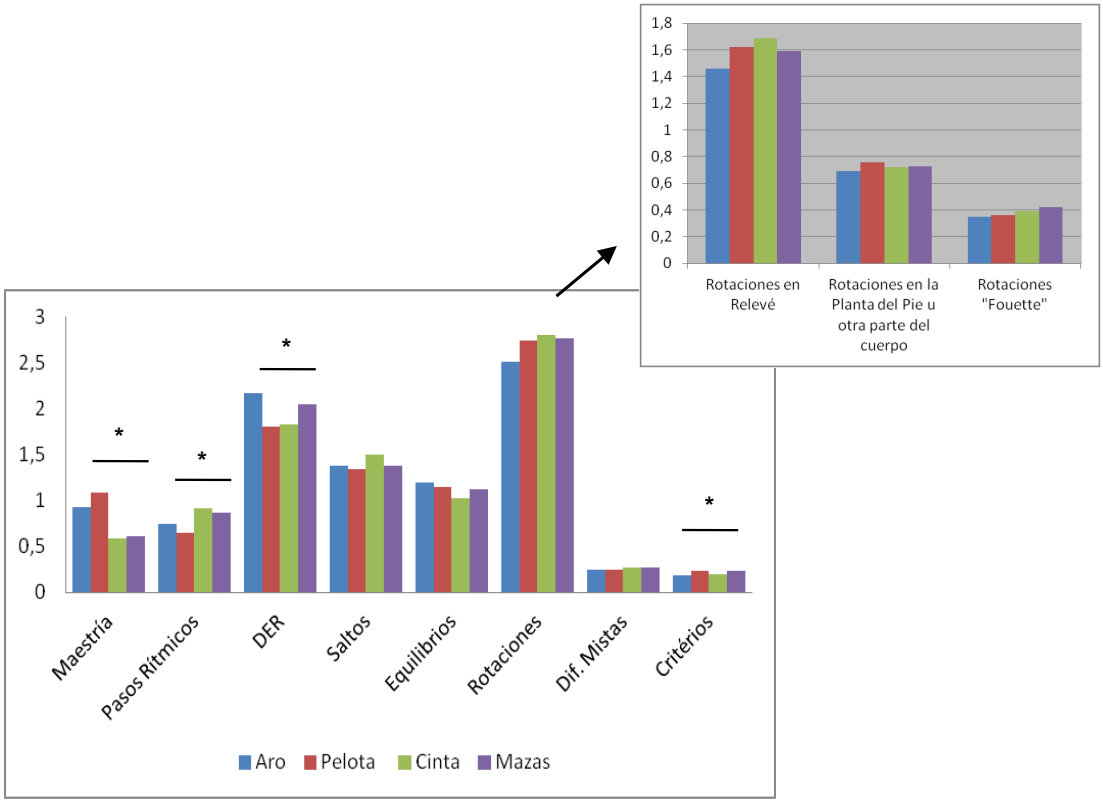


Figura 1. Media del Valor de los elementos de dificultad presentes en los ejercicios con los 4 aparatos (Kruskal Wallis, *p<0.05).

Tabla 1. Estadística Descriptiva y valores del test Kruskal Wallis con *p<0.05 para el Valor de los elementos de dificultad presentes en los ejercicios con los 4 aparatos.

	Aro (n=72)		**Pelota** (n=72)		**Cinta** (n=72)		**Mazas** (n=72)		Kruskal-wallis
Valor de la Dificultad	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**P=**
Maestria	0.93±0.56	0-2.6	1.09±0.61	0-3.4	0.58±0.44	0-1.8	0.61±0.41	0-2	0.000*
Pasos Rítmicos	0.74±0.27	0.3-1.5	0.65±0.22	0.3-1.2	0.91±0.25	0.3-1.8	0.87±0.25	0.3-1.5	0.000*
DER	2.17±0.30	1.5-2.8	1.80±0.24	1.1-2.4	1.83±0.27	1.0-2.3	2.04±0.31	1.2-2.7	0.000*
Critérios	0.18±0.14	0-0.6	0.23±0.12	0-0.6	0.20±0.14	0-0.7	0.23±0.13	0-0.7	0.007*

En cuanto al número de elementos de dificultad presentes en los ejercicios de Aro, Pelota, Cinta y Mazas, destaca el elevado número de elementos con Maestría. Si analizamos la presencia de los tres grupos de dificultad corporal (saltos, equilibrios y giro) destaca el mayor número de dificultades de Giro en Aro, Pelota, Cinta y Mazas (3,19 ± 0, 59; 3:31 ± 0,64; 3:39 ± 0,59, 3,35 ± 0,60), respectivamente, y el menor número de

dificultades de equilibrio en Aro, Pelota, Cinta y Mazas (02:56 ± 0,99; 0,97 ± 2:50; 2:22 ± 0,98; 2:44 ± 1:04) respectivamente. (Figura 2).

Entre los diferentes aparatos observamos diferencias significativas sólo en el número de elementos con Maestría, los Pasos Rítmicos y Criterios asociados a la dificultad (Tabla 2), especialmente en el número de elementos de Maestría entre la Cinta y el Aro (p = 0:00), Cinta y Pelota (p = 0:00), Mazas y Aro (p = 0,01) y Mazas y Pelota (p = 0:00); y en el número de elementos de Pasos Rítmicos entre Pelota y Mazas (p = 0:00), Pelota y Cinta (p = 0,00), Aro y Mazas (p = 0,01) y Aro y Cinta (p = 0:00); y en el número de elementos de los Criterios asociados con dificultad entre Aro y Mazas (p = 0,05) y Aro y Pelota (p = 0,01) (Figura 2).

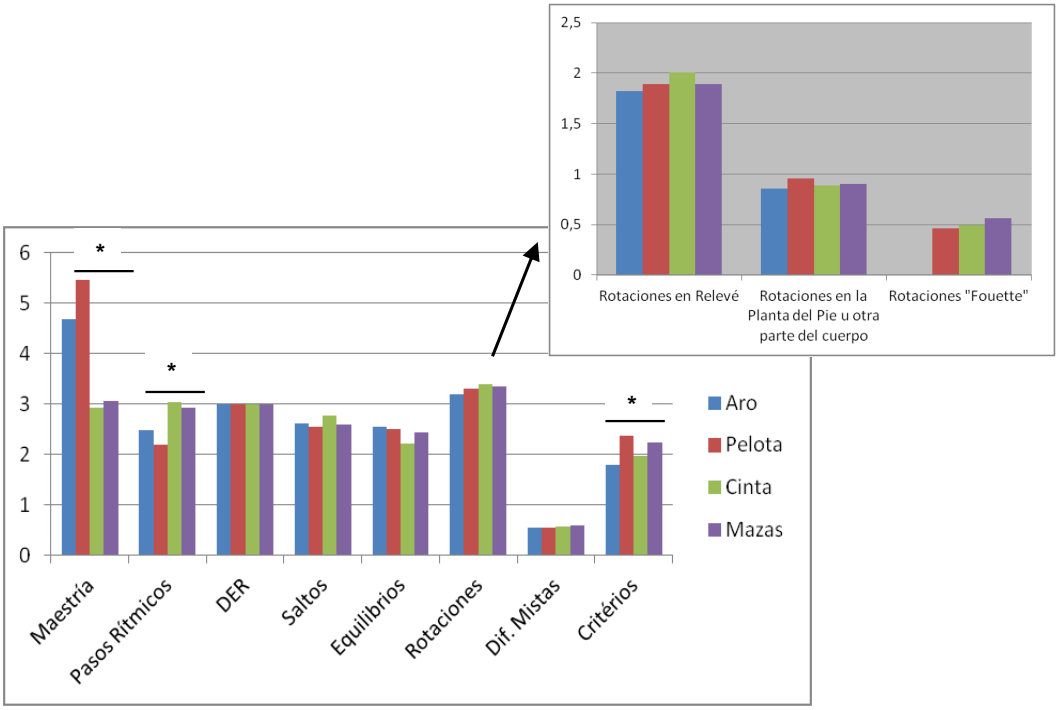


Figura 2. Media del número de elementos de dificultad presentes en los ejercicios con los 4 aparatos. (Kruskal Wallis, *p<0.05)

Tabla 2. Estadística descriptiva y valores del test Kruskal Wallis *p<0.05 para el número de elementos de dificultad presentes en los ejercicios, de los 4 aparatos.

	Aro (n=72)		**Pelota** (n=72)		**Cinta** (n=72)		**Mazas** (n=72)		Kruskal-wallis
Número Dificultad	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**Media+sd**	**Min-Max**	**P=**
Maestria	4.68±2.84	0-13	5.47±3.09	0-17	2.92±2.23	0-9	3.07±2.09	0-10	0.000*
Pasos Rítmicos	2.49±0.91	1-5	2.19±0.74	1-4	3.04±0.86	1-6	2.92±0.85	1-5	0.000*
Critérios	1.79±1.48	0-6	2.38±1.29	0-6	1.97±1.35	0-6	2.25±1.23	0-5	0.005*

Dificultades de Saltos

Observando la figura 3, vemos los Saltos de valor 0.5 son los más representados en los ejercicios en todos los aparatos. Le siguen los Saltos de valor 0,7 y 0,6. Los Saltos de menos valor son los menos representados en los ejercicios individuales en todos los aparatos. También se encontró un ligero predominio de los Saltos de valor 0,5 en Pelota, de valor 0,6 en Cinta y de valor 0,6 y 0,7 en Aro.

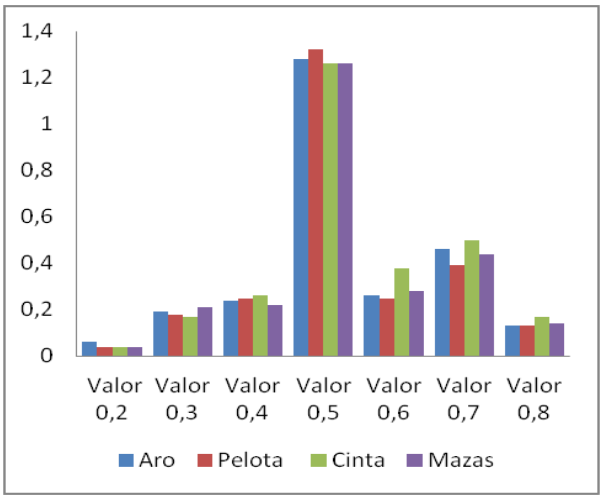


Figura 3. Número de elementos de dificultad de Salto presentes en los ejercicios con los 4 aparatos.

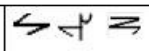
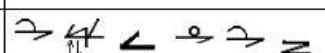
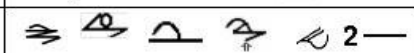
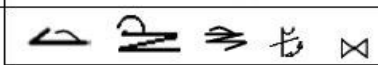
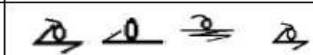
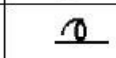
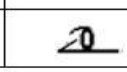
La Tabla 3 muestra que no hay diferencias significativas (prueba de Kruskal-Wallis) en la frecuencia de aparición de los diferentes Saltos, cuando se realizan con diferentes aparatos. Sin embargo, analizando los diferentes Saltos entre sí, los saltos de valor 0,5 registran diferencias significativas (prueba de Friedman), respecto a los Saltos con otros valores.

Tabla 3. Estadística Descriptiva, valores de test Kruskal Wallis y test Friedman, para el número de los elementos de dificultad de Salto presentes en los ejercicios de los 4 aparatos.

Saltos	**Aro** (n=72)	**Pelota** (n=72)	**Cinta** (n=72)	**Mazas** (n=72)	Kruskal-wallis test
	Media+sd	*Media+sd*	*Media+sd*	*Media+sd*	*P=*
Valor 0.2	0.06±0.28	0.04±0.26	0.04±0.26	0.04±0.26	0.954
Valor 0.3	0.19±0.46	0.18±0.45	0.17±0.44	0.21±0.47	0.922
Valor 0.4	0.24±0.45	0.25±0.49	0.26±0.50	0.22±0.45	0.979
Valor 0.5	1.28±0.80	1.32±0.74	1.26±0.88	1.26±0.73	0.967
Valor 0.6	0.26±0.47	0.25±0.49	0.38±0.59	0.28±0.51	0.543
Valor 0.7	0.46±0.52	0.39±0.51	0.50±0.58	0.44±0.52	0.717
Valor 0.8	0.13±0.33	0.13±0.37	0.17±0.44	0.14±0.38	0.964
Teste Friedman	0.000	0.000	0.000	0.000	

Analizando los diferentes tipos de Saltos, destaca dentro de los saltos de valor 0.5 con mayor frecuencia la “Zancada girando" (31%) y en los saltos de valor 0.7 la "Zancada girando con flexión del tronco" (17%). También podemos observar que en la cinta es donde se produce mayor utilización de saltos (26,4%), aunque la diferencia entre los aparatos no es significativa

Tabla 4. Número y Porcentaje de elementos de dificultad de Salto, presentes en los ejercicios, de los 4 aparatos.

Saltos		Aro	Pelota	Cinta	Mazas	Total	%
Valor 0,2		<5	<5	<5	<5	13	**1,7%**
Valor 0,3		<10	<10	<10	<10	54	**7,1%**
Valor 0,4		<10	<10	<10	<10	70	**9,2%**
Valor 0,5		<5	<5	<5	<5	18	**2,4%**
		27	32	29	26	114	**15,0%**
		62	56	58	61	237	**31,2%**
Valor 0,6		<10	<10	<10	<10	84	**11,1%**
Valor 0,7		33	28	36	32	129	**17,0%**
Valor 0,8		9	9	12	10	40	**5,3%**
		188	184	200	187	759	**100%**
		24,8%	**24,2%**	**26,4%**	**24,6%**	**100%**	

Dificultades de Equilibrios

Observando la figura 4, encontramos que los Equilibrios de valor de 0.5 son los más representados en los ejercicios en todos los aparatos. Los siguientes son, en orden descendente, los Equilibrios de valor de 0,4 y 0,3. También se encontró un ligero predominio de los Equilibrios de valor 0.5 en Aro, de valor 0.4 en Pelota y de valor 0,3 en las Mazas.

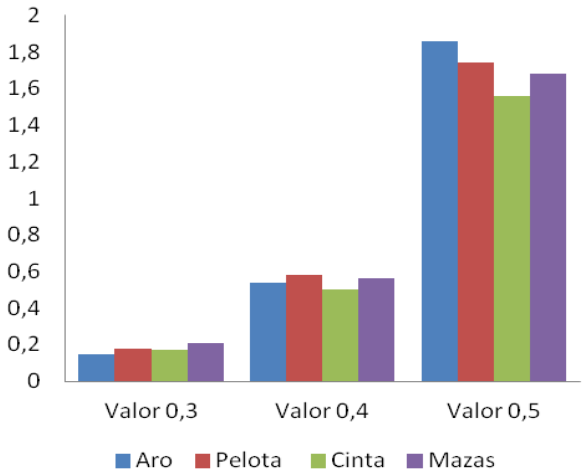


Figura 4. Número de elementos de dificultades de Equilibrio presentes en los ejercicios con los 4 aparatos.

En la tabla 4 podemos ver que no se registran significativa (prueba de Kruskal-Wallis) en la frecuencia de los distintos tipos de Equilibrios en los diferentes aparatos. Sin embargo, los Equilibrios de valor 0,5 presentan diferencias estadísticamente significativas en comparación con los equilibrios de otros valores (test de Friedman).

Tabla 5. Estadística Descriptiva y valores del test Kruskal Wallis y test de Friedman, para número de los elementos de dificultad de Equilibrio presentes en los ejercicios con los 4 aparatos.

Equilíbrio	**Aro** (n=72)	**Pelota** (n=72)	**Cinta** (n=72)	**Mazas** (n=72)	Kruskal-wallis test
	Media+sd	*Media+sd*	*Media+sd*	*Media+sd*	*P=*
Valor 0.3	0.15±0.39	0.18±0.42	0.17±0.41	0.21±0.44	0.835
Valor 0.4	0.54±0.71	0.58±0.70	0.50±0.73	0,56±0,78	0.83
Valor 0.5	1.86±1.21	1.74±1.10	1.56±1.12	1.68±1.24	0.41
Teste Friedman	0.000	0.000	0.000	0.000	

Si analizamos los diferentes tipos de equilibrios, destacan con mayor frecuencia los equilibrios de valor de 0,5, especialmente el "Equilibrio con la pierna en la vertical y con el tronco horizontal" al lateral, por delante y por detrás (18,1%, 16,4% y 16,3% respectivamente). Podemos observar que es en Aro donde se produce con mayor utilización de los Equilibrios (26,3%).

Tabla 6. Número y Porcentaje de elementos de dificultad de Equilibrio, presentes en los ejercicios con los 4 aparatos.

Equilibrio		Aro	Pelota	Cinta	Mazas	TOTAL	%
Valor 0,3		<10	<10	<10	<10	51	7,3
Valor 0,4		<10	<10	<10	<10	119	17
		<20	<20	<10	<10	38	5,4
Valor 0,5		<10	<10	<10	<10	40	5,7
		29	24	30	31	114	16,3
		32	34	31	30	127	18,1
		29	31	25	30	115	16,4
		34	29	15	18	96	13,8
		184	180	160	176	700	100%
		26,3%	25,7%	22,8%	25,2%	100%	

Dificultades de Giros

En cuanto a la Figura 5, vemos que los Giros de valor de 0,3 en relevé y rotaciones en la planta del pie u otra parte del cuerpo, son los más utilizados en los ejercicios en todos los aparatos. Sin embargo, vemos un ligero predominio de los Giros de el valor 0,3 en las Mazas y los giros en la planta del pie u otra parte del cuerpo de valor 0,4 en Pelota.

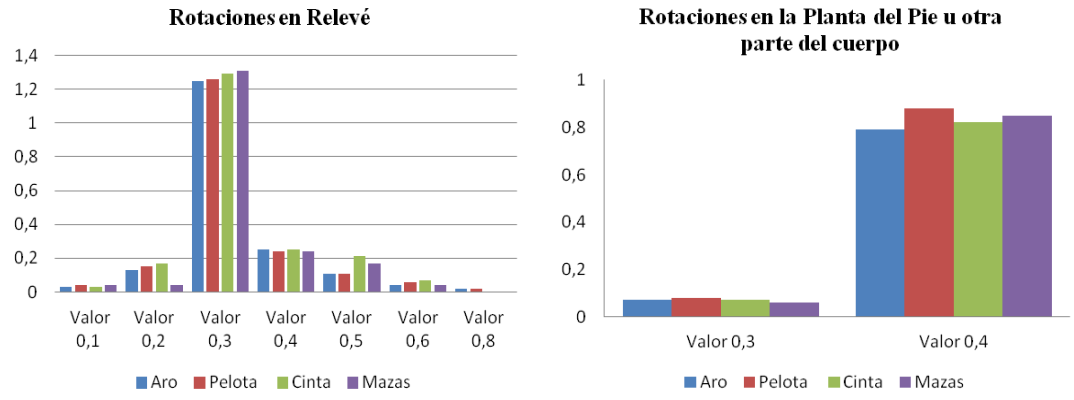


Figura 5. Número de elementos de dificultad de Rotación presentes en los ejercicios con los 4 aparatos.

En las Tablas 7 e 8 se puede observar que no hay diferencias significativas (prueba de Kruskal-Wallis) en la frecuencia de los diferentes Giros ya sea en relevé ya sea en la planta del pie o en otra parte del cuerpo en los diferentes aparatos. Sin embargo, en el análisis de los diferentes Giros en relevé entre sí, los Giros de valor 0,3 presentan diferencias significativas (prueba de Friedman) para todos los otros tipos. Lo mismo se verifica en los Giros en la planta del pie o de otra parte del cuerpo de valor 0,4.

Tabla 7. Estadística Descriptiva y valores del test Kruskal Wallis y test Friedman, para el número de los elementos de dificultad de Rotación en Relevé presentes en los ejercicios con los 4 aparatos.

Rotaciones en Relevé					
	Aro (n=72)	**Pelota** (n=72)	**Cinta** (n=72)	**Mazas** (n=72)	Kruskal-wallis test
	Media+sd	*Media+sd*	*Media+sd*	*Media+sd*	*P=*
Valor 0.1	0.03±0.16	0.04±0.20	0.03±0.16	0.04±0.20	0.938
Valor 0.2	0.13±0.37	0.15±0.43	0.17±0.41	0.04±0.20	0.761
Valor 0.3	1.25±0.78	1.26±0.82	1,29±0.81	1.31±0.74	0.932
Valor 0.4	0.25±0.46	0.24±0.42	0.25±0.49	0.24±0.48	0.986
Valor 0.5	0.11±0.35	0.11±0.39	0.21±0.47	0.17±0.41	0.274
Valor 0.6	0.04±0.20	0.06±0.23	0.07±0.25	0.04±0.20	0.856
Valor 0.8	0.01±0.11	0.01±0.11	0.00±0.00	0.00±0.00	0.571
Teste Friedman	0.000	0.000	0.000	0.000	

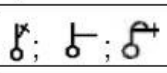
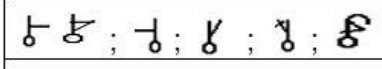
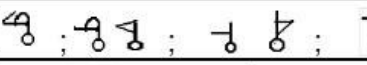
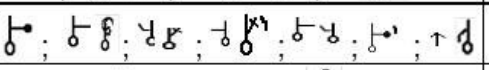
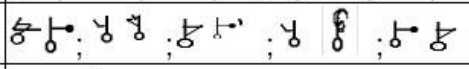
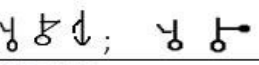
Tabla 8. Estadística Descriptiva y valores del test Kruskal Wallis y test Friedman, para el número de los elementos de dificultad de Rotación en la planta del pie u otra parte del cuerpo presentes en los ejercicios con los 4 aparatos.

Rotaciones en la Planta del Pie u otra parte del cuerpo					
	Aro (n=72)	**Pelota** (n=72)	**Cinta** (n=72)	**Mazas** (n=72)	Kruskal-wallis test
	Media+sd	*Media+sd*	*Media+sd*	*Media+sd*	*P=*
Valor 0,3	0,07±0,25	0,08±0,27	0,07±0,25	0,06±0,23	0,934
Valor 0,4	0,79±0,47	0,88±0,58	0,82±0,51	0,85±0,46	0,834
Teste Friedman	0,000	0,000	0,000	0,000	

En el análisis de la frecuencia de los diferentes tipos de Giros, (tabla 9) observamos que dentro de los Giros de valor de 0,3 en relevé destaca el "Giro en attitude" y el "Giro con la pierna libre por encima de la horizontal con ayuda" (18,6%, 14,9% respectivamente), y los resultados de los Giros en la planta del pie de valor 0,4 destaca el "Giro en penché" con el 27% de las frecuencias.

Hemos observado que la cinta presenta en una mayor utilización de los Giros es (25,9%), aunque la diferencia entre los aparatos no es significativa.

Tabla 9. Número y Porcentaje de elementos de dificultad de Giro, presentes en los ejercicios con los 4 aparatos.

Rotaciones en Relevé		Aro	Pelota	Cinta	Mazas	TOTAL	%
Valor 0,1		<5	<5	<5	<5	10	**1,2%**
Valor0,2		<10	<10	<10	<10	39	**4,8%**
Valor o,3		<10	<10	<10	<10	98	**12,1%**
		33	27	30	30	120	**14,9%**
		35	38	39	38	150	**18,6%**
Valor 0,4		<10	<10	<10	<10	70	**8,7%**
Valor 0,5		<5	<5	<5	<5	65	**8,1%**
Valor 0,6		<5	<5	<5	<5	15	**1,9%**
valor 0,8		1	1	0	0	2	**0,2%**
	SUB- TOTAL	*134*	*144*	*151*	*140*	*569*	**70,5%**
Rotaciones en la Planta del Pie u otra parte del cuerpo							
Valor 0,3		<5	<5	<5	<5	20	**2,5%**
Valor 0,4		54	54	53	57	218	**27,0%**
	SUB- TOTAL	*59*	*60*	*58*	*61*	*238*	**29,5%**
	TOTAL	**193**	**204**	**209**	**201**	**807**	**100%**
		23,9%	**25,3%**	**25,9%**	**24,9%**	**100%**	

3.2.5. Discusión

Hemos llevado a cabo un análisis de los resultados (cantidad, número y tipo) en los aparatos Aro, Pelota, Mazas y Cinta en dos formas de análisis: (1) el análisis integral de la composición de los ejercicios; (2) el análisis por grupos de diferentes elementos de dificultad utilizado en los ejercicios.

En el análisis global del valor de la dificultad en la composición de los ejercicios, nos encontramos con un valor promedio muy similar en todas las gimnastas y en todos los aparatos, alrededor de 9,3 puntos, cerca de la puntuación máxima posible, que es de 10 puntos. Este resultado sugiere un alto nivel de excelencia global en todas las gimnastas, pero de hecho, como según se muestra en la pantalla de los resultados finales (FIG, 2013), sólo las 8 mejores gimnastas puede llegar a él. Por lo tanto, se observa que el valor del ejercicio propuesto por el entrenador en la ficha de la competición es ambicioso y no refleja la mayoría de las veces la capacidad real de rendimiento de la gimnasta. Además, el hecho de que el valor propuesto sea similar en los diferentes aparatos, nos permite ver que las gimnastas no presentan diferencias en los aparatos, ni en la técnica corporal ni en la técnica de aparato, lo que en cierta medida refleja una depreciación de esta, ya que como las exigencias de la técnica corporal del CP son igualen en todos los aparatos, la técnica del aparato podría o incluso debería reflejar diferencias. Señalar también que el grupo de los Giros y los DER en conjunto representan alrededor del 50% del valor de la dificultad de la composición total, en todos los aparatos. Este porcentaje se debe al aumento de los criterios asociados con estos dos grupos de dificultad que aumentan el grado de complejidad de su ejecución (CP 2012). Encontramos aquí un importante cambio en el COP en este ciclo olímpico, ya que en los ciclos olímpicos anteriores el mayor valor de la composición era proporcionado por grupo de los Saltos (Caburrasi, 2003; Ávila, 2011). Según Leandro (2015), se trata de una modernización en GR ya que supone un aumento de la complejidad en la ejecución de los ejercicios. La fuerte interacción entre la gimnasta y el aparato aumentó el grado de dificultad de los ejercicios caracterizando el desarrollo de la GR (Lebre, 2011). Por otro lado, la preferencia dada a estos dos grupos dificultad puede afectar a la variedad y la diversidad en la composición de ejercicios, características necesarias para enriquecer la composición de ejercicios (Balcells, 2009; Leandro, 2015). Este aumento de valor tan significativo en sólo dos grupos de dificultad, puede llevar a preferirlos en detrimento de los otros grupos. La limitada variedad en la elección de las dificultades en la composición de los ejercicios los vuelve aburridos y compromete su valor artístico (Ávila, 2012a). También el hecho de que este resultado sea similar en todos los aparatos, refleja la necesidad de incluir habilidades complejas en los ejercicios para buscar altas puntuaciones (Massida, 2012), sin importar el aparato

en cuestión. Las dificultades de Equilibrio son la categoría menos utilizada en los ejercicios en todos los aparatos. Este resultado puede estar relacionado con el tiempo de ejecución, ya que son esencialmente elementos de dificultad estática (Gateva, 2015) y también porque están menos valorados por el código (0,50 es el máximo posible para el Equilibrio). (Agopyan, 2014) obtuvo resultados similares para el ciclo olímpico anterior. Los elementos de dificultad de Maestría y Pasos rítmicos (características específicas de la GR e introducidas por primera vez en este ciclo Olímpico en los requisitos de composición), comparativamente, están lejos de alcanzar los valores obtenidos por los elementos de dificultad corporal en todos los aparatos. Estos grupos son menos valoradas por el CP. Por lo tanto, habrá que aumentar, la complejidad de la ejecución, así como su valor para que estos grupos tan característico de GR sean evidentes en la composición (Leandro, 2015). Reseñamos también el hecho de que en todos los aparatos es menor el número de Pasos Rítmicos que el de Maestrías con el aparato, a pesar de que tienen un grado de complejidad en su ejecución técnica inferior. En nuestra opinión, puede tener que ver con el tiempo 8 segundos, requerido para realizar los pasos rítmicos que puede condicionar la elección de los entrenadores y la preferencia de éstos, ya que sólo se dispone de un minuto y medio para gestionar el cumplimiento de todos los requisitos de la composición. También comprobamos que el tipo de elementos dificultad realizados dentro de cada grupo corporal, Saltos, Equilibrios y Giros se repiten varias veces en las coreografías, sin que se registren diferencias significativas entre los aparatos. Los elementos de dificultad identificados con mayor frecuencia de ejecución son el “Giro en attitude”, “Giro en penché”, el “Equilibrio con la pierna libre en la vertical y el tronco en posición horizontal sin ayuda” y en los saltos la "Zancada girando" y "Zancada girando con flexión del tronco". Sin embargo, cuando se comparan los elementos de dificultad dentro de cada grupo, nos encontramos con una diferencia estadísticamente significativa de los elementos anteriores, respecto a todos los otros elementos de dificultad. Estos resultados muestran que a pesar de variar el aparato que la gimnasta manipula, y el trabajo específico propio de este, la composición de los ejercicios no se caracteriza por ser única, ni diversa y creativa, en cuanto a la técnica corporal. Para promover la modalidad bien como deporte de alto rendimiento o bien como espectáculo deportivo es absolutamente imperativo que cuando el espectador mire el ejercicios con un aparato determinado, no sea similar al de otro

aparato sino una sorpresa y nueva emoción, por asistir a una combinación de elementos de técnica corporal y técnica de aparatos característica y única de aquel aparato, que según Pelin, (2013) refleje la espectacularidad de la coreografía. Esta combinación de especificidad corporal con la singularidad de cada aparato debe justificar la competición de la misma gimnasta con cada aparato, mostrando su versatilidad, y distinguiéndola de las demás. Cuando analizamos las diferencias entre los aparatos por grupo de dificultad no observamos diferencias significativas en la presencia de los grupos de dificultad de base corporal (Saltos, Equilibrios y Giros) en la composición de los ejercicios tanto en el análisis cuantitativo como en el análisis cualitativo. Apenas encontramos diferencias significativas en el número y el valor de la Maestría, de los Pasos rítmicos y de los Criterios asociados a la dificultad, que se utilizan en cada aparato. En nuestra investigación en bases de datos EBSCO, Scopus y PubMed, celebrada en noviembre 2015 encontramos estudios de análisis de contenido técnico relativos a saltos, giros y equilibrios. (Trifunov y Slobodanka de 2013, y (Agopyan, 2014)). No se encontraron estudios relativos al examen de otros grupos de dificultad (de maestría, pasos rítmicos, DER y Criterios asociados con dificultad (ondas y elementos preacrobáticos), por lo que no es posible hacer un análisis comparativo de los resultados. Hacemos por lo tanto, un análisis basado en las características técnicas de este deporte y de los aparatos en particular. El grupo de las Maestría se realiza con más frecuencia en los ejercicios de Pelota y Aro. Especulamos que esto ocurre tanto por las características físicas de estos aparatos, como por las características técnicas específicas del mismo, en particular por los grupos técnicos básicos cada uno. Teniendo en cuenta los requisitos generales establecidos por el CP “la Maestría del aparato es la combinación de elementos no ordinarios de aparato” (FIG, 2012). Nos parece que la forma "redonda" y consistente (no se deforma), así como los elementos fundamentales de rodamiento sobre el cuerpo y el suelo característicos de la Pelota y el Aro permite más y mejores combinaciones de los criterios corporales definidos para la realización de la Maestría. No es así con la Cinta que es un aparato suave y deformable con una alta complejidad de ejecución en el manejo y menor número de elementos fundamentales. También en las Mazas el hecho de que se manejen en realidad 2 aparatos que deben de trabajar simultáneamente y no sólo uno como en los otros casos, puede condicionar la combinación de sus elementos técnicos específicos con los criterios corporales. Curiosamente, por el contrario, es

precisamente en estos dos ejercicios con las Mazas y la Cinta que se producen más pasos rítmicos con una diferencia significativa a los ejercicios de Aro y Pelota. Teniendo en cuenta el análisis de estos dos grupos, Maestría y Pasos rítmicos en paralelo, parece evidente que esto sucede por la estrategia del entrenador de una perspectiva de compensación, tanto por el valor, como por la gestión del tiempo. Hecho este análisis, parece que siendo estos dos grupos fundamentales en el aumento de la parte artística, especifica del deporte, la danza, el ritmo y la manipulación del aparato deberían de ser privilegiados en la composición de los ejercicios con todos los aparatos, lo que sólo podría suceder si los requisitos de composición del CP fueran diferentes. Los Criterios asociados a la dificultad, las ondas y Elementos preacrobáticos, ocurren con más frecuencia en los ejercicios de Aro, lo que en nuestra opinión era previsible. El COP exige que la realización de estos elementos se acompañe por trabajo del aparato, y esta más facilitada en el Aro ya que su estructura física permite la inclusión de pasos por dentro del aparato, un elemento técnico que le distingue de otros aparatos y de fácil combinación con ondas y preacrobacias . También en los DER, encontramos diferencias significativas en los diferentes aparatos. Hay un mayor valor de los DER en los ejercicios de Aro y Mazas, resultados previsibles ya que el número de posibles criterios para añadir cuerpo es mayor en estos aparatos por definición, el código puntuación FIG, 2012).

3.2.6. Conclusiones

Los resultados obtenidos nos indican que las gimnastas del Campeonato Mundial en KIEV 2013 utilizan dificultades físicas similares en la composición de sus ejercicios en los diferentes aparatos. Los elementos de dificultad que se realizan con mayor frecuencia son el “Giro en attitude”, “Giro en penché”, “Equilibrio con la pierna libre en la vertical y el tronco horizontalmente sin ayuda” y el salto “Zancada girando”, con una diferencia estadísticamente significativa respecto a todos los demás tipos elementos de dificultad. Los grupos de Dificultad con mayor valor en la composición son los DER y los Giros y representan un aumento significativo en el coeficiente de dificultad de los ejercicios. El Equilibrio es el grupo de dificultades corporales menos utilizado en todos los aparatos. Las principales diferencias estadísticamente significativas en la composición de los ejercicios en los distintos aparatos, fueron los siguientes: (i) Número

de Maestrías, Pasos rítmicos y Criterios asociados con dificultad; (ii) El valor de las Maestría, Pasos rítmicos, Criterios asociados con dificultad y DER.

En resumen, los resultados muestran características estructurales en la composición ejercicios muy similares en los diferentes aparatos respecto a los aspectos de la técnica corporal que determinan el mayor porcentaje en la composición, Los ejercicios de los diferentes aparatos se distinguen en los grupos más relacionados con la parte artística (Maestría, Pasos rítmicos y DER), principalmente debido a las características específicas de cada aparato y no por requisitos del CP. Los Pasos rítmicos y las Maestrías son los grupos menos valorados por lo que en el resultado final no se refleja una verdadera diferencia en el valor de la composición en cada aparato. Este estudio proporciona información actualizada sobre el contenido técnico de los ejercicios individuales de gimnasia rítmica de élite, para ser considerado respecto a: (i) la posibilidad de modificar el presente Código de Puntuación, sobre todo en la definición de las exigencias de composición que favorezcan la variedad y diversidad y fomentar el valor artístico y la unidad técnica del aparato; (ii) el proceso de formación de valor y el perfil de desempeño de GR gimnastas de elite.

3.2.7. Referencias

Agopyan, A. (2014). Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. *Journal of Scientific Research, 19*(12), 1554-1565.

Ávila-Carvalho, L., Klentrou, P., Palomero, M. d. L., & Lebre, E. (2012) Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. *The Open Sports Sciences Journal,* 5, 146-153.

Ávila-Carvalho, L., Klentrou, P., & Lebre, E. (2012b). Handling, Throws, Catches and Collaborations in Elite Group Rhythmic Gymnastics. *Science of Gymnastics Journal, 4 (3), 37-47*.

Ávila-Carvalho, L., Leandro, C., & Lebre, E. (2011). 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), *Book of abstracts of the 16th Annual Congress of the European College of Sport Science.* Liverpool, UK. 579-580.

Balcells, M., Martín, C., & Anguera, M. (2009). Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión Corporal y Danza Contact-Improvisatio. Apunts educación física y deportes.Ciencias aplicadas a la actividad física y el deporte, 14-23.

Caburrasi, E.F., Santana, M.V. (2003). Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002. Granada2003 Available from: http://www.efdeportes.com/efd65/grd.htm.

FIG. (2012). Code of Points for Rhythmic Gymnastics Competitions. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

FIG. (2013). Gymnastics Results. Available at: http/www.gymnasticsresults.com

Gateva, M., Gospodarski, N., Treneva, V., Avramov, D., Ivanov, N., Andonov, K. (2015). Comparison Between The Static Balance Of Practitioners From Different Sports and Non-Athletes. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science*. p. 569. Malmo, Sweden.

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). What Do Rhythmic Gymnastics Judges Think About Their Code Of Points?. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science* (pp. 569). Malmo, Sweden.

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015b). Analysis of the Technical Content of Elite Rhythmic Gymnastics Individual Routines of different levels. Science of Gymnastics Journal. (Accepted for publication 15 november, 2015).

Lebre, E. (2011). Technical principles for the new framework. Crossroads to the Future. International Federation of Gymnastics Scientific [Press release].

Massidda, M. & Calò, M.C. (2012). Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, Journal of Sports Sciences, (30)13, 1415-1420

Pelin, R. A.(2013) Studies Regarding The Rhythmic Gymnastics From The Olympic Games. Sport & Society / Sport si Societate, (13), p61.

Sierra-Palmeiro, E., Fernandez-Villarino, M., Bobo-Arce, M. (2015). Estudio longitudinal (1997 - 2014) de la técnica de aparatos en al gimnasiaritmica individual de

élite. Editado por Universidade Lusófona de Lisboa, Livro de resumos do 1º Congresso Iberoamericano "Desporto, Educação, Atividade Física e Saúde". Lisboa, Portugal.

Trifunov, T., & Slobodanka, D. (2013). The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. *Physical Culture, 67*(2), 120-129.

Vitrichenko, N., Klentrou, N., Gorbulina, N., Della Chiaie, D. & Fink, H. (2011). Groups. In FIG (Ed.), *Rhythmic Gymnastics. Technical Manual. Level 3.* 3-55. Lousanne: FIG Academy.

Wang,M., Lu,M., Sun, X. (2013). Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. Journal of Physical Education / Tiyu Xuekan, (20) 5, p117.

4. Discussion and Conclusions

4. Discussion and conclusions

Throuh final considerations we try to descant on a multifaced, progressive and integrated vision of the process that leads to success in rhythmic gymnastics elite competition. To potentiate the success in competition it is important to understand deeply the variables involved (Ferreirinha, 2011).

We organize the final considerations according to the two phases that guided our work. The 1st phase was devoted to the study of the variables of evaluation in RG and the 2nd phase was devoted to the analysis of the performances in excellence of individual routines of RG, in particular to the technical content of difficulty elements of the routines of world reference gymnasts.

It was our intention to reflect on the two phases of the process of competition that seem to us determinant and constitute the key points to improve the quality of the high level competitions in the future, on the way to success and excellence (Cuk, 2012).

Thus, before suggesting future changes, it is important to understand how it works in the present, finding out what should be changed and what should be kept.

From this perspective, to accomplish this purpose we have structured the present dissertation in sixe separate papers.

The four first papers were dedicated to the study of performance evaluation variables in competition. The proposal to do this 4 studies follows a leading line that allows to do a progressive and evolutive analysis of the way to go, in order to reach success and excellence in the evaluation of the competition.

We define this leading line in:

i) Analyse the efficiency of the resources involved and the accuray in the judgement

ii) Charaterizing and contextualizing the resources involved, judges and evaluation instruments.

iii) Analyse the quality of the result of applying these resources.

iv) Consider the connection between the science and the practice in the analysis of the efective implications from the juding process to the training/competition process.

This way we started by doing a first and second studies dedicated to the characterization of the evaluators/judges as well as understanding and analysing their knowledge on a global perspective, about the present evaluation tool, the CoP (FIG2012), and if these efectively influence the practice of judging and the consequent performance of the judges, particularly in the evaluation of body difficulty elements.

In one study, we can verify that the majority of international judges has a high academic level, works as a RG coach and has a large experience in judgement. This type of background offer efficiency conditions to the judges, since they register quality judging values, when compared to quality values of judgement got by judges with less experience, as well as lower academic level (Leandro, Ávila-Carvalho, & Lebre, 2010).

The criteria that can distinguish a "specialist in the matter" goes from the connection of the specialist with the problem, the professional experience, the personal qualities or professional ability, to the guaranty of the quality of the answers and the skill of recognizing detailed information (Almenara, 2013).

Besides de characterization of the judges, we tried to know what they think about the CoP, once they are the ones who use it as an evaluation tool in RG routines. The judges are the ones using it as an evaluation tool thus their opinions represent a reference value to be considered in its elaboration. We conclude that the instrument of evaluation used right now is not yet ideal to absolutely assure the validity and reliability in RG judgment. The judges manifested different opinions about the objectivity in the evaluation of the Difficulty, Execution, as well as the different parameters of evaluation in artistic faults. They highlighted the Mastery, Dance Steps and DER, in Difficulty and the Artistic Faults (mainly "Unity Composition" and "Music/Movement") in Execution, as being the ones with most complexity in evaluation when considered the objectivity. They suggested changes in the evaluation criteria of these Evaluation classe`s, in order to become more precise in the final evaluation. Finally the judges stated that they would like to have in the CoP some rules to limit the repetition of difficulties in the different apparatus routines, in order to improve the diversity and variety in RG routines, promoting an enrichment of the sport.

In the other study, we tried to analyse, the factors that may be in the origin of the biggest evaluation constraints , such as the different levels of performance of gymansts and/or the different apparatus used, Hoop, Ball, Cubbs and Ribbon. We

verified that the four judges of difficulty panel did not agree in their evaluation in 40% of the difficulty elements presented in the difficulty forms. Regarding the final ranking of the gymnasts the agreement level is higher in the high level gymnasts and decrease in the low level gymnasts. The level of accuracy was lower in the difficulty elements which validation criteria depends not only from difficulty criteria but also from execution criteria.

The analysis by type of difficulty elements showed that for the Jumps and Balances the judges agreed on the evaluation of the elements which means an acceptable accuracy of judgement, but for the other types of elements the level of disagreement between the judges was significantly high to be an accurate judgement, where we highlight the Mastery and DER difficulty elements. This study confirms the results of the first study and provides updated information about the precision of difficulty judging in rhythmic gymnastics, to be considered in the definition of the evaluation criteria of the elements where we see the highest disagreement between judges.

After diagnosing constraints in the evaluation of individual RG routines in elite competition, we tried to understand if these are reflected in the result of the final score, which led us to the third study, the analysis of the reliability and validity of Difficulty judging on Rhythmic Gymnastics at different levels of performance. Overall, the results suggest that the reliability of the judgment in RG is satisfactory in the first and third parts of the ranking, once the Cronbach's alpha is above 0.90, minima of item total correlations and the ICC of average scores are above 0.80. Overall, for the World Championships analyzed, regarding the final ranking of the gymnasts, the indices of consistency are satisfactory in both high and low level gymnasts.

The level of the consistency indices was lower in the 2nd part of the ranking. When trying to explain the inferior reliability results for medium level gymnasts it is valuable to inspect the between-judge correlation matrix, as many of the reliability measures of judges' performances are based on Pearson's correlations. We can identify several judges (without highlighting any over the others) whose correlation coefficients are below 0.5 in all apparatus.

The validity in our analysis was assessed through systematic bias in judging, considered as repetitive under- or over-estimation of particular judges. When looking at

the results as a whole, systematic bias in individual judge's scores and judges' panels was modest or poor in the 2nd part of the ranking. It is obvious that the quality of judging differs when evaluating different levels of gymnasts' performances. Our results show that the biases in the judgment of rhythmic gymnastics competition routines are not so much due to the performance of specific judges, but more to the differences in the level of performance of the gymnasts at the same competition.

To further clarify the factors contributing to the observed, we can speculate that these differences are perhaps a source of additional variability in the judge's scores and that part of the problem may originate in the judging rules (Code of Points) that are not well defined to evaluate the gymnasts whose execution is not clear and precise. The apparatus used by the gymnast does not seem to be a cause of variability in the judging since we found equivalent values of mean deviations from final D scores for all apparatus, in each of the parts of the ranking.

It's important here to analyse the link between the results of studies 1 and 3. In both studies, we understand that the judgement of the medium performance level gymnasts's routines is more complex once it implies problems of precision, validity and reability. It's also important to highlight that these two papers together lead us to two very important reflexions. The first one, in which the international judges of RG have high performance levels in a global perspective and when the evaluation criteria leave no doubts, and the second in which the judges give each routine the same final score having validated elements of different difficulties.

The 4th study tries to analyse the effects of the judges evaluation in the Departure scores, as a result for the training process by the coaches, considering the real success in competition.

According to the results we can see that there are big differences between the scores proposed by the coaches in the competition forms and the difficulty elements that judges could identify, which increase as the gymnasts go lower in the ranking, going from almost more than 1 point in the gymnasts placed in the top ranking to almost 4 points for the gymnasts placed in the bottom of the ranking, with a statistically significant difference between the 3 level of gymnasts. We can state that the difficulty value of the proposed routine by the coach in the competition form is ambitious and does not reflect the performance capacity of the gymnast.

To help explain these results, we tried to understand if these differences between the Departure Difficulty Score and Final Difficulty Score could be identified in the different types of difficulty elements. We found out that they exist in all difficulty elements, being significantly higher in the difficulty elements Rotations, DER and Mastery and almost residual in the difficulty elements Dance Steps for the gymnasts in the three groups of the ranking. Analysing the elements first mentioned, these results may suggest the coaches and the judges understand the CoP rules differently and/or the coaches do not have a real perception of the performance capacity of their gymnasts, in these types of difficulty elements.

To sum up, taking into consideration the specific objectives for the first phase of our study we can conclude that in relation to the:

1st objective: To analyse the accuracy of the judges' evaluation of the difficulty elements performed by the RG high level individual gymnasts in their competition routines.

- The judges' evaluation is not accurate in 40% of the analysed cases and the evaluation is more complex in the routines of medium level gymnasts and the difficulty elements mastery, DER and rotations.

2nd objective: To analyse the evaluation tool (CoP) from the perspective of the judges.

- The instrument of evaluation used right now is still not ideal to absolutely assure the validity and reliability in RG judgment.

3rd objective: To analyse the quality of judging within the group of judges who evaluated a high-level RG individual competition.

- The quality of judging differs when evaluating different levels of gymnasts' performances, and is more complex in the medium performance level gymnasts' routines.

4th objective: To analyse how judges and coaches understand the CoP requirements and the routine content in the success competition.

- The coaches and the judges understand the CoP rules differently and/or the coaches do not have a real perception of the performance capacity of their gymnasts.

In the 2nd phase of this dissertion contruction we did the last two papers, 5 and 6, dedicated to the study of excellency performance variables in individual RG routines, in high level competitive context. The analysis of difficulty elements was structured in 2 parts: body difficulties and apparatus difficulty and considering the type, number and level of the selected elements for the composition of the competition individual routines.

The analysis of the competition routines in RG allows us to identify and characterize the competitive context, creating data bases in order to identify the most influencing performance indicators and the tendencies in the development of RG. Allows us also to support for the evolution of the coaching process, defining performance profiles for individual gymnasts and ranking performances.

The presented studies on both paper 5, relative to the technical content of elite rhythmic gymnasts routines, and paper 6 relative to the quantitative and qualitative analysis of the rhythmic gymnastics indlividual routines in the different apparatus (variety and diversity) allowed us to consider that there is a common pattern in the composition of the competition routines in high performance levels. The RG gymnasts who competed at the 2013 World Championships used in their routines very similar difficulty elements with limited variety. In these two papers 5 and 6, independently of the performance level of the gymnast and/or the apparatus used, we can see that generally the composition of individual routines uses more frequently body difficulties base on the rotation “attitude”, rotation with “free leg in ring in back with help”, “rotation in penché”; balance “side scale with split, without help” and balance “back scale leg high up”; jump “jeté with turn” and “jeté with a turn with back bend”. The highest valued elements are DER and rotations and they represent 50% of the total value of the routine. These groups showed an important contribution to the final D score. This increment of such significative value in only two difficulty groups may lead to the preference of these instead of other groups. The balances were the less used difficulty body group.

In paper 5 we can see that the routines had differences in the composition pattern between the gymnasts according to the their final ranking in the following items: (i) the number of rotations of flat foot or other part of the body, “fouetté” rotations and MixDif; (ii) the value of jumps, rotations, DER and MixDif. Concerning the

dance steps and mastery, no differences were found between the routines of gymnasts place in the three parts of the ranking.

In paper 6, the main diferences statistically significant in the composition pattern of the routines in different apparatus were: (i) number of masterys, dance steps and criteria associated with the difficulty (ii) value masterys, dance steps, criteria associated with the difficulty and DER. The body difficulty groups (jumps and balances), do not show a significant diference in the composition of the routines, on both qualitative and quantitaive analysis. The group of masterys is performed more in Ball and Hoop routines. We can speculate that such happens because of both the physical characteristics of these apparatus and the specific technical characteristics of the apparatus, in particular of the fundamental groups in each of them. Also in DER, the results show higher values in the routines of Hoop and Cubbs, an awaited result once the number of possible criteria to add to the number of body rotations is higher in these two apparatus, by definition of the CoP (FIG, 2012).

We found in these two studies complementing results, which demonstrates very similar structural characteristics of routines compositions on both different gymnasts levels and different apparatus routines, concerning body technic which determines the highest percentage of composition value. The routines of different apparatus show more diferences in the groups more directly related to the artistic part (masterys, dance steps and DER), essentialy because of the specific characteristics of each apparatus and not because of specific demands of CoP or the level of performance of the gymnast. The dance steps and masterys are less valued difficulty groups so in the final result, it does not correspond to a real difference in the composition value of each apparatus.

These results show that although the apparatus changes, the composition of the routines is not characterized by being unique, with diversity and creativity, concerning body technique. It is important to give value to the specific work in each apparatus interacting with variated and diverse body difficulty, in order to enrich and unify each coreography. To promote the sport as a high level sport as well as a spectacle in the sports phenomenon is absolutely imperative that when the spectator is watching the routine of a certain apparatus, it is not expectable a similar routine as for another appartus, but a surprise and new emotion, by seeing a combination of body tecnique elements and specific and unique tecniques for that specific apparatus. This

combination of body specificities with the singularity of each of the apparatus should justify the competition of the same gymnast in the different apparatus, revealing her versitility, distinguishing her from the rest.

Taking into consideration the specific objectives for the second phase of our study we can conclude that in relation to the:

5th objective: To analyse the content of RG individual routines regarding the body elements and apparatus elements, according to the gymnasts' performance in competition.

- Concerning the variables in the performance of individual routines RG competition, we highlight that DER and rotations represent more than 50% of the total value of the composition and that can be considered the presence of a common pattern in the composition of competition routines, on gymnasts of different performance levels as well as routines with different apparatus, by the usage of very similar difficulty elements with limited variety

Finally, by identifying some gaps between these two variables directly implied in the competition success, the papers developed in this dissertation have a potential impact on this debate. Our findings may be useful from an applied perspective in order to optimize the environments in which the RG gymnasts try to reach the highest levels of performance to find highest levels of performance.

In the end, having the notion that our main findings are based on informations retrospectively recalled, and recognizing that research and practice should be as close as possible, this dissertation supports the value, of an analysis of the competitive context to the highest level, in the perspective of evaluation and performance variables to be considered to benefit gymnastes' progression and the possible modifications of the Code of Points, in particular for the definition of the composition requirements in order to have higher level of variety and diversity in the routines, and increment the artistic and technical value of the apparatus toward success and excellence.

5. Limitations and Implications of the studies

5. Limitations and implications of the studies

Implications

The present study is pioneer in its field since it is the first one to analyze the judgement process with official validated data from a competition at the highest performance level (World Championships). The gymnasts' evaluation in competition was made with the official difficulty forms used, completed and signed by the judges during the competition. The access to these information was only possible due to the fact that the International Gymnastics Federation (FIG) has a strong commitment to improve the evaluation process and results in gymnastics competitions, following the Olympic Agenda 2020 20+20 Recommendations, (IOC, 2014).

Concerning the "*Evaluation of the judging process in Rhythmic Gymnastics*" the results obtained provide fundamental and updated information related to:

- the profile of high level international judges;
- the standing point of the International judges concerning the structure and the applicability of the rules stated in the CoP;
- the quality and preciseness of judging the Difficulty Component in rhythmic gymnastics;
- the particularities of the judging process in different levels of performance and in exercises performed with different apparatus, hoop, ball, clubs and ribbon;
- the specific parameters that are more complex to evaluate;

The results and the information they provide are crucial and should be considered to:

- Create data bases for the identification of the international RG judges educational, professional and expertise ideal profile as a guarantee of high level of performance quality in the judgment process.
- Contribute for a better definition of the evaluation criteria for some of the technical element groups.

- Improve changes in the CoP, mainly regarding the definition of the evaluation criteria in order to reach higher levels of reliability and validity in judgment.
- Drive the attention to medium level gymnast performance evaluation, since the evaluation directives included in the CoP are not well precise for the gymnasts at this level.
- Orientate the training process to maximize the performance capacity of the gymnasts on some groups of elements, allowing an effective success in competition.

Concerning the "*Analysis of the content of high level competition routines"*, the results found in this study can deliver relevant information concerning the planning of the training process. Although the competition routines are all distinct, the number and type of elements performed at the competitive routine is a valuable indicator to plan the training load, and to establish the correlation of volume and intensity according to the competitive period.

The analysis of the content of competitive routines provides us with updated information related to:

- The main characteristics of the competition routines content (type elements chosen and its organization in the routine).
- Differences and similarities of the competition routines performed by gymnasts of different technical levels.
- Difficulty value of the competition routines performed by gymnasts of different technical levels.

These results are very useful and should be considered to:

- Guide and to enhance the choices and work strategies of the coaches in the planning and conducting of the training process.
- Optimize the training process to achieve the highest-level of performance in the individual gymnasts.

- Improve the diversity and variety of technical elements in RG competition routines, promoting an enrichment of the sport.
- Contribute to establish the profile of the RG competition routines with different type of apparatus and in different levels of performance.
- Promote, with a scientific base, consistent changes in the CoP. In this sense, some of the results obtained in this study have been already used by the FIG Scientific Commission Working Group to implement new directives in the Code of Points 2017/2020 that will model the sport and its evaluation in the new Olympic Cycle.

Limitations

The main limitation of this study is the fact that it is a transversal study and not a longitudinal one. It would be ideal that we could reproduce the study in all the World Championships of the 2012/2016 Olympic Cycle, but we only had permission from FIG to access the completion forms used in Kiev 2013.

We identify also some limitations with the procedure and the questioners delivered to the International judges. There were some judges that did not answer. We could not accede to judges from some countries (e.g. North Korea) because of language and/or cultural barrier. Nevertheless, the percentage of judges with a valid brevet in the present cycle, that answered is a representative sample of the total population.

Suggestions fot future studies

Considering the objectives, conclusions and limitations of this thesis, new research proposals arise to enrich the knowledge of this RG, its evaluations and training processes.

The judgement evaluation is a very little explored field and there are enormous possibilities for future research, like:

- Comparative study of the understanding of the CoP rules by judges from different world regions.
- Study the stability of the judging results of the same exercises with different judging panels.
- Study the difficulty elements in the routines that contribute to the highest scores.

6. References

6. References

1. Agopyan, A. (2014). Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. *Journal of Scientific Research, 19*(12), 1554-1565.
2. Almenara, J. (2013). La utilización del juicio de experto para la evaluación de tic: el coeficiente de competencia experta. Universidad de Sevilla, Bordón, 65 (2), 25-38.
3. Askew, S. (2002). Feedback for learning. Journal of eduction for teaching, 28, 83-90.
4. Arkaev, L.I. & Suchilin, N.G.(2004). How to create champions - the theory and methodology of training top-class gymnasts. Oxford: Meyer & Meyer Sport.
5. Ávila-Carvalho, L., Palomero, M. d. L., & Lebre, E. (2009c). Difficulty score in Group Rhythmic Gymnastics. Portimão 2007/2008 World Cup Series. *Palestrica Mileniului III. Civilizatie si sport, Anual X, 3*(37), 261-267.
6. Ávila-Carvalho, L., Leandro, C., & Lebre, E. (2011). 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), *Book of abstracts of the 16th Annual Congress of the European College of Sport Science.* Liverpool, UK, 579-580.
7. Ávila-Carvalho, L., Palomero, M. L., & Lebre, E. (2011b). Estudio del valor artístico de los ejercicios de conjunto de Gimnasia Rítmica de la Copa del Mundo de Portimão 2007 y 2008. Apunts. Educación Física y Deportes, 1.er trimestre, 103, 68-75.
8. Ávila-Carvalho, L., Klentrou, P., Palomero, M. L., & Lebre, E. (2012). Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. *The Open Sports Sciences Journal,* 5, 146-153.
9. Ávila-Carvalho, L., Klentrou, P., & Lebre, E. (2012b). Handling, Throws, Catches and Collaborations in Elite Group Rhythmic Gymnastics. *Science of Gymnastics Journal, 4 (3), 37-47*.
10. Balcells, M., Martín, C., & Anguera, M. (2009). Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión

Corporal y Danza Contact-Improvisatio. Apunts educación física y deportes.Ciencias aplicadas a la actividad física y el deporte, 14-23.

11. Bartolomeis, F. (1999). Avaliação e Orientação: Objectivos, instrumentos e métodos. Lisboa: Livros Horizonte.
12. Boen, F. B., Karen, H., Yves, V. A., Jos, F., & Tim, S. (2008). Open feedback in gymnastic judging causes conformity bias based on informational influencing. Journal of Sports Sciences, 26(6), 621–628.
13. Bobo, M. (2002). El juicio deportivo en Gimnasia Rítmica. Una propuesta de evaluación basada en indicadores de rendimiento. (PHD Thesis), Universidad da Coruña, Instituto Nacional de Educación Física de Galicia, Coruña.
14. Bobo, M., Sierra, E. (2003). Estudio de las repercusiones de los cambios de código de puntuación en la composición de los ejercicios de gimnasia rítmica en la técnica corporal. Available from: http://www.cienciadeporte.com/congreso/04%20val/pdf/p3.pdf.
15. Bobo, M., Sierra, E. (1998). Una nueva propuesta de dificultades corporales en gimnasia rítmica deportiva. Libro de resúmenes del VI Congreso de Educación Física e Ciencias do Deporte dos Países de Lingua Portuguesa Deporte e Humanismo en clave de Futuro.
16. Bučar, M., Cuk, I., Pajek, J., Karacsony, I., & Leskosek, B. (2011). Reliability and validity of judging in women's artistic gymnastics at University Games 2009. European Journal of Sport Science, 12, 207-215.
17. Bučar, M., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? Journal of Human Kinetics, 37, 173-181.
18. Bučar, M., Kovač, M., Pajek, J., Leskošek, B. (2014). The Judging of artistry components in female gymnastics: a cause for concern? Science of Gymnastics Journal, 6 (3), 5-12.
19. Caburrasi, E.F., Santana, M.V. (2003). Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002. Granada2003 Available from: http://www.efdeportes.com/efd65/grd.htm.

20. Catteeuw, P., Helsen, W., Gilis, B., Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing. Journal Sport Science, 27, 1125-1136.

21. Čuk, I., & Karacsony. (2004). Vault: methods, ideas, curiosities, history. Ljubljana: STD. Sangvinck.

22. Čuk, I., Fink, H., & Leskošek, B. (2012). Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. Science of Gymnastics Journal, 4, 73-82.

23. Dallas, G., & Kirialanis, P. (2010). Judges' evaluation of routines in men's artistic gymnastics. Science Gymnastics Journal, 2, 49-58.

24. Debien, P., Noce, F., Debien, J., Costa, V. (2014). Stress in rhythmic gymnastics refereeing: a systematic review. Revista da Educação Física/UEM, 25 (3), 489-500.

25. Fernandez-Villarino, M., Bobo-Arce, M., Sierra-Palmeiro, E. (2013). Practical Skills of Rhythmic Gymnastics Judges. Journal of Human Kinetics, 39, 243-249.

26. Ferreirinha, J., Carvalho, J. (2012). Tendencies and deviations of judging in gymnastics. Revista ENGym, 1, 2-3.

27. Ferreirinha, J., Carvalho, J., Côrte-Real, C., Silva, A. (2011). Evolução do Valor Real de Dificuldade dos Exercícios de Paralelas Assimétricas de Ginástas de Elite nos Ultimos Ciclos Olímpicos. In FGP (Ed.), Da Prática à Ciência. Artigos do 2º e 3º Congresso da Federação de Ginástica de Portugal, 71-78.

28. FIG-International Gymnastics Federation. (2012). Code of Points for Rhythmic Gymnastics Competitions, 2012; Available at: http://www.fig-gymnastics.com/site/page/view?id=472

29. FIG-International Gymnastics Federation. (2013). Gymnastics Results, 2013; Available at: http/www.gymnasticsresults.com

30. Findlay, C., & Ste-Marie, M. (2004). A Reputation Bias in Figure Skating Judging. Journal of Sport & Exercise Psychology, 26 (1), 154-166.

31. Flessas, K., Mylonas, G. & al. (2015). Judging the Judges' Performance in Rhythmic Gymnastics. Medicine & Science in Sports & Exercise. American College of Sports Medicine, 47(3), 640-648.

32. Gateva, M. (2014). Investigation of the effect of the training load on the athletes in rhythmic and aesthetic group gymnastics during the preparation period, Research in Kinesiology, 4, 40-44.

33. Gateva, M., Gospodarski, N., Treneva, V., Avramov, D., Ivanov, N., Andonov, K. (2015). Comparison Between The Static Balance Of Practitioners From Different Sports and Non-Athletes. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science*, Malmo, Sweden, 569-569.

34. Gambarelli, G., Laquinta, G., & Piazza, M. (2012). Anti-collusion indices and averages for the evaluation of performances and judges. Journal of Sports Sciences, 30(4), 411-417.

35. Guardo, M. (2004). Hasta una teoria del arbitrage deportivo. In http:/www.efdeportes.com/ Rev. Digital. Buenos Aires, 68-68.

36. Heinen, T., Vinken, P., & Velentzas, K. (2012). Judging Performance in Gymnastics: A Matter of Motor or Visual Experience? Science of Gymnastics Journal, 4, 63 – 72.

37. Hökelmann, A., Breitkreutz, T., & Liviotti, G. (2012). *Changes in performance structure during group competitions in rhythmic gymnastics*. Edited by Prof. Derek M. Peters & Dr. Peter G. O'Donoghue, *Book of abstracts* of the World Congress of Performance Analysis of Sport IX University of Worcester, UK, 99-99.

38. IOC-International Olympic Committee. (2014). Olympic-Agenda-2020; Available at: https://stillmed.olympic.org/media/Document%20Library/OlympicOrg/Documents/Olympic-Agenda-2020/Olympic-Agenda-2020-20-20-Recommendations.pdf#_ga=1.113878351.978678035.1471607188

39. Johansson, A. (2010). Deciding who is the best. Validity issues in selections and judgements in elite sport. Academic dissertation, Faculty of Social Sciences, Umeå University, Sweden.

40. Kirkpatrick, J., & Hawk, L. (2006). Curricula and evaluation: Maximizing results. Measuring and Evaluating. Available from EBSCO.

41. Leandro, C. (2009). Avaliação de Juízes de Ginástica Rítmica. (Master Thesis), Porto University, Porto.

42. Leandro, C., Ávila-Carvalho, L., & Lebre, E. (2010). The avaluation of the performance of Rhythmic Gymnastics` Judges. Palestrica of the Third Millennium Civilization & Sport, 11(3), 202-206.

43. Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines. Science of Gymnastics Journal, 7(3), 81-93.

44. Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). What Do Rhythmic Gymnastics Judges Think About Their Code Of Points?. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science* ,Malmo, Sweden, 569-569.

45. Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2016). Technical Content of Elite Rhythmic Gymnastics. Science of Gymnastics Journal, 8(1), 85-96.

46. Lebre, E. (2011). Technical principles for the new framework. Crossroads to the Future [Press release]

47. Leskošek, B., Čuk, I., Bučar-Pajek, M. (2015). Trends in E and D scores and their influence on final results of male gymnasts at European Championships 2005–2011. Science of Gymnastics Journal, 5 (1), 29-38.

48. Liviotti, G., Hökelmann, A. (2012). Which Quantifiable Performance Parameter(s) Determined the Medals Winners at the World Championship 2011 In Rhythmic Gymnastics – Group Competition?. In Rsupesy & T.(Ed.), Book of abstracts of the VI International Scientific Conference of Students and Young Scientists, Modern University Sport Science, Moscow, 43-43.

49. Liu, X.X., Kuang, L., (2001) Review of evolvement course of international evaluation rules in rhythmic gymnastics and its effects on technique development. Journal of Beijing Sport University, (24)3, 412-415.

50. Mark, M., & Shotland, R. (1987). Multi Methods in Program Evaluation. New Directions For Program Evaluation, Londres: Jossey – Bass Inc, 35.

51. Martin, S. (2006). Caracteristicas deseables para el arbitraje y el juicio deportivo em judo. Revista Iberoamericana de Psicologia del ejercicio y el Deporte, 1, 27-40.

52. Massidda, M. & Calò, M.C. (2012). Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, Journal of Sports Sciences, (30)13, 1415-1420.

53. Morrow, J., Jackson, A., G-Disch, J., Mood, D. (1995). Measurement and Evaluation in Human Performance, Library of congress cataloging - in - Publication data. Mystat. ISBN: 0-87322-731-X (BooK), USA.

54. Palomero, M. L. (1996). Hacia una objetivación del Código Internacional de Gimnasia Ritmica Deportiva.Barcelona.

55. Plessner, H. (2005). Positive and negative effects of prior knowledge on referee decisions in sports. T. Betsch, Haberstroh, S., (Ed.) The routines of decision making, 311–324.

56. Plessner, H., & Schallies, E. (2005). Judging the Cross on Rings: A Matter of Achieving Shape Constancy. Applied Cognitive Psychology, 19, 1145-1145.

57. Pelin, R. A. (2013). Studies Regarding The Rhythmic Gymnastics From The Olympic Games. Sport & Society / Sport si Societate, (13), 61-61.

58. Popovic´, R. (2000). International bias detected in judging rhythmic gymnastics competition at Sydney 2000 Olympic Games. Facta Universitatis (Series: Physical Education and Sport), 1, 1-13.

59. Sands, w.A., Caine, D.J., Bornes, J. (2003). Scientific aspects of women´s gymnastics, Basel: S. Karger A.G., 45.

60. Sierra-Palmeiro, E., Fernandez-Villarino, M., Bobo-Arce, M. (2015). Estudio longitudinal (1997 - 2014) de la técnica de aparatos en al gimnasiaritmica individual de élite. Editado por Universidade Lusófona de Lisboa, Livro de resumos do 1º Congresso Iberoamericano "Desporto, Educação, Atividade Física e Saúde". Lisboa, Portugal.

61. Simões, G. (2000). A avaliação do desempenho Docente. Lisboa: Texto Editora.

62. Ste-Marie, D., & Lee, T. D. (1991). Prior processing effect on gymnastic judging. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 126-136.

63. Ste-Marie, D. M., Valiquette, S. M., & Taylor, G. (2001). Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. Research Quarterly for Exercise and Sport, 72(4), 420-426.

64. Tamir, P. (1998). Assessment and Evaluation in Science Education . Opportunities to Learn and Outcomes International Hand book of science Education: Dordrecht: Kluwer Academic Publishers, 761-789.

65. Trifunov, T., Slobodanka, D. (2013). The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. Physical Culture, 67(2), 120-129.

66. Vitrichenko, N., Klentrou, N., Gorbulina, N., Della-Chiaie, D., Fink, H. (2011). Groups. In FIG (Ed.), Rhythmic Gymnastics. Technical Manual. Level 3, Lousanne: FIG Academy, 3-55.

67. Wang, M, Lu, M., Sun, X. (2013). Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. Journal of Physical Education / Tiyu Xuekan, (20) 5,117-117.

7. Appendix

Appendix 1. Published paper 1: Accuracy in judgment the difficulty score in elite rhythmic gymnastics individual routines.

ACCURACY IN JUDGMENT THE DIFFICULTY SCORE IN ELITE RHYTHMIC GYMNASTICS INDIVIDUAL ROUTINES

Catarina Leandro [1,2], Lurdes Ávila-Carvalho[2], Elena Sierra-Palmeiro[2], Marta Bobo[2]

[1] Faculty of Psychology, Education and Sport, University Lusófona of Porto, Porto, Portugal
[2] Faculty of Sport Science and Physical Education, University of Coruña, Coruña, Spain

Original article

Abstract

The main goal of this study was to analyse the accuracy in judging the Difficulty score in the Rhythmic Gymnastics Kiev World Championship 2013. The accuracy was determined analysing the judges' agreement on the evaluation of the routines difficulty elements. 1152 difficulty forms concerning 288 individual routines were analysed - 4 forms per routine, 1 per judge. To allow the comparison between gymnasts with different levels the individual routines were clustered into 3 subgroups according to their final ranking competition. Body difficulty elements were organized, according to the composition requirements stated in the RG Code of Points (FIG, 2012). Non-parametric tests - Cochran's Q and Chi-Square Tests were applied to determine whether there were significant differences between groups. As main results we can point out that in general the judges did not agree on difficulty evaluation in 40% of the elements. The level of accuracy was lower in the second part of the ranking, and in the Mastery and DER difficulty elements.

Keywods: *Evaluation, accuracy, judge, rhythmic gymnastics.*

INTRODUCTION

Rhythmic Gymnastics (RG) is characterized by the high level of difficulty of the body elements and apparatus handling, combining esthetical and artistic components. This complexity increases the difficulty of the judgment and its accuracy mainly in high level performances. The requirements are quantitative (amount and variety of body and apparatus movements) and qualitative (degree of difficulty and quality in performance) and they are evaluated by the judges according to the rules and evaluation criteria stated in RG Code of points (Bobo, 2002).

The Body and apparatus movements are grouped according to the type of skills, the level of difficulty and the complexity of the movements (Lebre, 2011). The main groups considered in the routines evaluation are: Jumps, Balances and Rotations, Mixed difficulties, additional criteria for the body movements - waves and pre-acrobatics,

Dance Steps, Mastery (special apparatus handling) and Dynamic Elements with Rotation and throw (DER).

In competition the performance is evaluated by 2 panels of judges: the difficulty (D) jury that judges the routines content and the execution (E) jury to evaluate the quality of the routines. The gymnasts present in each competition a difficulty form with all difficulties listed. Each judge must confirm the difficulty elements performed by the gymnast and cross out those that are not correctly performed or not performed at all (FIG, 2012). The final D score is the average of two intermediate scores. When the score become published on the screens, the judges can compare the final score to their own scores. Therefore, the judges score independently although there's still some feedback (Bucar, Cuk, Pajek, Kovac, & Leskosek, 2013).

In previous studies was noted that judging is not only a matter of identifying the sports performance. There are also various facts, identified in the literature, having an influence on the several stages of processing information in gymnastics judgment (Leandro, 2009).

Findlay and Ste-Marie (2004) found out that the were the judges tend to judge better the gymnast higher qualified in previous competitions, concluding that the reputation of the gymnasts have influence on the judging. The judge's experience has been also described as influencing the quality of judgment. Leandro, Ávila-Carvalho, and Lebre (2010) and Ste-Marie, Valiquette, and Taylor (2001) found that the more experienced judges had better perception and anticipation of the elements and there for, were better evaluators. Other factors, as the memorizing capacity (Ste-Marie, Valiquette, and Taylor, 2001), and the tendency to adapt their scores to those given by the judges of the same panel (Boen, Karen, Yves, Jos, and Tim, 2008) were also described. The observation angle (Plessner and Schallies, 2005) and the judges with experience as gymnasts (Heinen, Vinken & Velentzas, 2012) were also described as factors that can influence in the judges accuracy.

Besides these factors, is also relevant to know whether the factors related to the sport specificity as the structure/organization of the Code of Points, the evaluation criteria defined by the sports authorities has an influence (positive or/and negative) on the judge's performance and consequently on the gymnasts final scores.

Rhythmic gymnastics has been experiencing a constant and outstanding evolution in its' technic for the last few years because of the evolution of the Code of Points (Palomero, 1996). The evaluation of the gymnasts is made by a collective observation of judges that should be objective. However, this evaluation is not yet exact, probably due to huge amount of evaluation criteria defined for each difficulty element. This can be verified by the differences registered between the judges of the same panel when the evaluate the same routine. This fact is wellknown in the sport but not yet studied. The majority of studies available deal with the analysis of the technical content of exercises or with the final scores given at the end of each exercise. We could not find any study dealing with the analysis of the difficulty evaluation, element per element, trying to see if the final score of each judge are the product of the validation of the same difficulty elements.

Under this subject, the most relevant studies we found are Palomero (1996) and Bobo (2002), in which both the authors present a new proposal for the scoring, based in performance indicators. Čuk, Fink, & Leskošek (2012) studied the way the different type of final score calculation can change the gymnasts final ranking. Gambarelli, Laquinta & Piazza (2012) developed a formula to avoid pre-agreements between judges. They proposed that the score from the judge of the same country of the gymnast should not enter in the calculation for the gymnast final score. Furthermore, they consider that this would be a factor of guarantee of higher reliability of the final score.

Some of the studies demonstrate that the structure of the Code of Points itself holds decisive influence in scoring gymnasts. In this way is very important to suggest alternative evaluation tools that respect the principles of evaluation (objectivity, validity, reliability, discriminating power and practical utility) and allow a balanced appreciation of the different dimensions of the sport, in either aspects of quality or quantity in the performance of gymnasts (Bobo, 2002).

On the other side, the permanent changes in the Code of Points may cause a lack of understanding of the rules, which lead to a need of evaluation of judging instrument itself (Kirkpatrick & Hawk, 2006). Mark & Shotland (1987) remarked, any evaluation model has to be based on a group of principles, axioms and postulates that must be feasible. To have a Code of Points with an extremely complex model of evaluation that does not work when it has to be used, must be avoid.

According to Bartolomeis (1999) it is not possible to see everything at the same time. The essential point is that the evaluation instrument evaluates what it is supposed to evaluate. For Tamir (1998) the evaluation criteria used should be tested in both validity (precision) and reliability (internal consistency).

We could not find any study based on the analysis of the judges' activity based on the using of the difficulty forms during the competition, making this study a pioneer in this field.

Thus, before suggesting future changes, it is important to understand how it works in the present, finding out what should be changed and what should be kept. According the pyramidal structure of the evaluation process (Figure 1) we established the goal of the study.

The goal of this study was to analyse the accuracy in judging Difficulty in the Kiev World Championship 2013, trying to learn if the 4 difficulty judges evaluate in the same way the difficulty elements on the D form (agreement between the 4 judges). This accuracy was studied for each element declared in the difficulty form trying to understand if the perception of the validation criteria for each elements is similar for all judges. The final difficulty score given by each judge to the same gymnast were very similar, but, with this study, we will analyse if the judges arrived to the final score validating the same elements or validating different elements.

Figure 1. Pyramidal structure for analysis of the evaluation process.

After analysing the date in a global way, we will study the level of agreement between the judges concerning the validation of the difficulty elements according to: (1) the position of the gymnast on the final ranking (1st part, 2nd part and 3rd part), (2) the routine apparatus (hoop, ball, clubs and ribbon), and (3) the type of difficulty element.

METHODS

Subjects and design

1152 difficulty forms concerning 288 individual routines were analysed (4 forms per routine, 1 per judge). The routines were performed by gymnasts from 45 different countries competing at Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013.

This study was done with the permission of the International Gymnastics Federation. Full blinding of the judges involved was undertaken.

All difficulty elements reported in the difficulty forms provided by the gymnasts at the competition were analysed. Each element was considered validate or not according the notes done by the judge on the form. For each element, we studied the cases of agreement when all 4 judges validate or not the difficulty element and the disagreement when at least one of the judge did not validate and the others consider the element correctly done.

The analyse was done considering the all sample, and the sample clustered into 3 subgroups according to gymnasts final ranking as follows: the first part of the ranking - the top 24 gymnasts, the second part of the ranking - 24 middle gymnasts and third part of the ranking – the 24 lower placed gymnasts on the ranking, to allow the comparison the agreement level of the judges when they evaluate gymnasts with different levels. Then, we studied the sample according to the apparatus used to perform the routine (hoop, ball, clubs and ribbon), and the type of difficulty element performed listed according to the composition requirements of the Code of Points (FIG, 2012), (Figure 2).

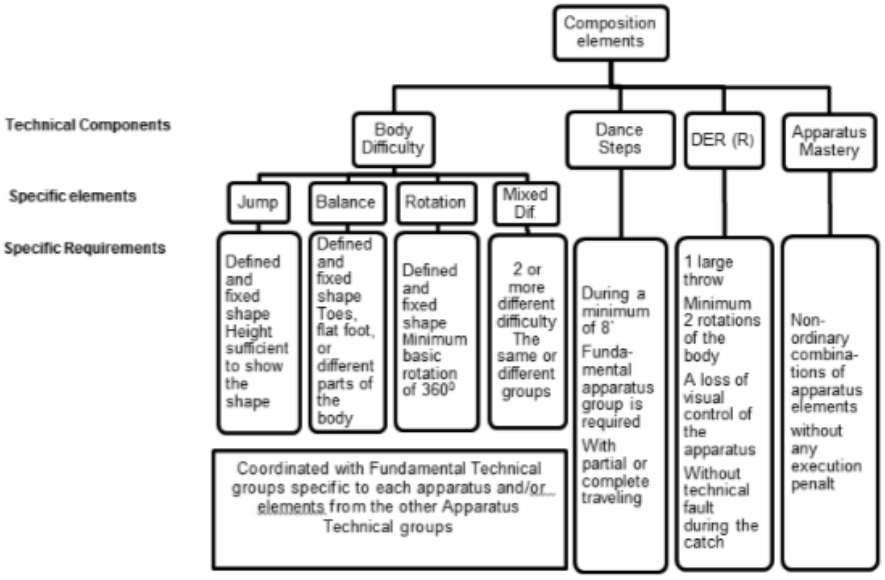


Figure 2. Technical Content of Rhythmic Gymnastics of Individual Gymnasts Routines (COP 2012/2016)

Statistical Analysis

For the statistical analysis we used the Statistical Package for the Social Sciences - Version 21.0 (SPSS 21.0, Chicago, USA) and Microsoft Office Excel 2010.

Non-parametric tests (Cochran's Q and Chi-Square Tests) were applied to determine if there were significant differences between groups. We use the Chi-square Tests for two independent samples to study the differences between two groups for each variable and the Cochran`s Q test to analyse when a set of K differs significantly. Significance level was set at $\alpha = 0.05$ (corresponding to a confidence level of 95%).

RESULTS

The forms were analysed first in a global way. For each difficulty element presented on the forms, the percentage of agreement between the 4 difficulty judges concerning the evaluation of the elements was determined. Then, the level of agreement on the elements evaluation was also calculated with the sample dived in 3 groups according to the final ranking of the gymnasts (Table 1).

The judges agreed on the evaluation of 60.0% of the difficulty elements presented on the difficulty forms. When we observe the results according to ranking of the gymnasts, is visible that higher the gymnast is placed in the ranking, higher is the agreement of the judges on the difficulty elements evaluation: 68.8% on the first part of the ranking, 56.1% on the 2nd part and 54.6% on the 3rd part. According to the results of the Chi-Square test, the differences between the cases of agreement and disagreement on the evaluation of the difficulty elements were statistically significant in all cases.

Table 1.
Level of agreement on the evaluation of the difficulty elements presented on the D Forms for all sample, and for the 3 groups according to the final ranking of the gymnasts.

	All Sample		1st part of the Ranking		2nd part of the Ranking		3rd part of the ranking	
	n	%	n	%	n	%	n	%
Not Agree	4871	40.0	1300	31.2	1836	43.9	1735	45.4
Agree	7294	60.0	2865	68.8	2343	56.1	2086	54.6

Chi-Square Test (Asymp.Sig.(2sided)) .000 * (*P<0.05)

Table 2.
Level of agreement on the evaluation of the difficulty elements presented on the D Forms according to the routine apparatus.

	Hoop		Ball		Ribbon		Clubs	
	n	%	n	%	n	%	n	%
Not Agree	1370	41.2	1129	37.3	1191	41.0	1244	40.6
Agree	1867	58.8	1894	62.7	1715	59.0	1818	59.4

Chi-Square Test (Asymp.Sig. (2sided)) .000 * (*P<0.05)

Table 3.
Results of the Cochran`s Q test comparing the results the agreement level for Hoop, Ball, Clubs and Ribbon routines.

	Hoop	Ball	Clubs	Ribbon
N	3174	3023	3062	2906
Cochran's Q	9.960	6.512	25.174	6.232
Sig.	**.018***	.090	**.000***	.099

(*P<0.05)

Table 4.
Results of the Cochran`s Q (C Q) test comparing the results the agreement level for Hoop, Ball, Clubs and Ribbon routines according to the final ranking of the gymnasts.

	Hoop			Ball			Clubs			Ribbon		
	1st part	2nd part	3rd part	1st part	2nd part	3rd part	1st part	2nd part	3rd part	1st part	2nd part	3rd part
N	1069	1078	1027	1044	1036	943	1050	1061	951	1002	1004	900
C Q	5.167	22.273	2.385	10.793	6.660	6.281	7.482	16.485	4.821	18.351	10.042	5.405
Sig.	.173	**.000***	.499	**.013***	.083	.095	.061	**.001***	.185	**.000***	**.019***	.145

(*P<0.05)

Table 5
Level of agreement on the evaluation of the difficulty elements presented on the D Forms according to the different type of elements.

	Not Agree		Agree	
	N	%	N	%
Mastery	726	62.5	436	37.5
Dance Steps	220	28.7	546	71.3
DER	1871	40.6	2735	59.4
Jumps	270	35.6	489	64.4
Balance	302	43.1	398	56.9
Rotations	1065	32.0	2263	68.0
Mixed Difficulties	93	38.3	150	61.7
Criteria assoc. to diff.	324	53.9	277	46.1

Chi-Square Test (Asymp.Sig.(2sided)) .000 * (*P<0.05)

Table 6
Results of the Cochran`s Q test comparing the results the agreement level for different groups of elements according to the final ranking of the gymnasts.

	1st part			2nd part			3rd part		
	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.
Jumps	257	1.227	.817	244	4.483	.208	258	2.92	.401
Balances	207	1.224	.785	238	5.89	.121	255	6.084	.106
Mastery	361	116.05	**.000***	394	46.744	**.000***	407	32.992	**.000***
DER	1607	62.548	**.000***	1567	8.492	**.036***	1432	17.251	**.001***
Dance Steps	260	8.12	**.047***	244	14.709	**.002***	262	2.121	.551
Rotations	1168	56.937	**.000***	1185	1.625	.652	975	4.288	.224
Mix. Diff.	108	10.553	**.015***	81	10.881	**.012***	54	8.937	**.030***
Criteria	197	12.425	**.005***	226	5.158	.164	178	3.774	.282

(*P<0.05)

Table 7.

Results of the Cochran`s Q test comparing the results the agreement level for different groups of rotations elements according to the final ranking of the gymnasts.

	1st part			2nd part			3rd part		
	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.	N	Cochran's Q	Sig.
RPIV Base	195	2,769	.586	167	2,780	.448	188	7,554	.051
RPIV Rotations	431	37,748	**.000***	346	1,213	.763	333	0,283	.969
RFF Base	99	2,314	.594	96	4,116	.295	65	7,627	**.050***
RFF Rotations	273	11.634	**.008***	198	3,915	.283	106	7,382	.060
RF	206	13,481	**.004***	378	7,774	**.050***	283	2,928	.419

Studying the difficulty forms according to the routine apparatus (Table 2) we observed that the range between the disagreement values for the elements evaluation in the 4 apparatus is not very wide (from 37.3% in ball to 41.2% in hoop). However, when we observed the results of the Chi-Square test we could verify that for all apparatus there were significant differences between the values of the agreement and the disagreement on the evaluation of the difficulty elements.

Comparing the data between apparatus through the Cochran`s Q test (Table 3) we could find that there is a significant difference between the values registered for Hoop and Clubs (p value 0.018 and 0.000 respectively), what showed that there was differences in judges agreement level on the elements evaluation for the different apparatus.

Continuing the analysis in each apparatus, we studied the lack of agreement between judges regarding the final ranking of the gymnasts.

The results of the Cochran`s Q test (Table 4) revealed that in Hoop, and Clubs the judges disagreed significantly only on evaluation the difficulty elements of the gymnasts ranked in the 2nd part of the final ranking; in Ball they disagree significantly on the gymnasts in the 1st part of the final ranking; and finally for Ribbon they disagree significantly on the 1st and 2nd part of the final ranking.

We studied the level of judges agreement on the difficulty elements considering the different group of elements described in the Code of Points (Table 5).

In the most part of the groups of elements the agreement percentage between the judges was higher than the disagreement percentage. Only for the evaluation of the Mastery group and the criteria associated to the difficulties (waves and acrobatic skills) the percentage of disagreement between the judges was higher than the agreement - 62.5% and 53.9% respectively for the agreement against 37.5% and 46.1% for the disagreement. Despite this remark, the results of the Chi-Square test the differences between the cases of agreement and disagreement on the evaluation of the difficulty elements were statistically significant in all cases.

The level of agreement between the judges evaluating the different groups of elements was, then, studied regarding the final ranking of the gymnasts (Table 6).

Observing the results we can see that for Jumps and Balances was not remarked a significant disagreement between judges on the evaluation of the elements performed by the gymnasts independently of their placement in the final ranking. For the Dance Steps, there was only a significant disagreement between the judges for the gymnasts placed in the first and second parts of the ranking. Regarding the Rotations and the Criteria associated to the difficulties the significant disagreement was registered only for the gymnasts placed on the first part of the ranking. When we observe the Table 6, we can see that there

are statistically significant differences for the Mastery elements, the DER elements and Mixed Difficulties in the 3 parts of the ranking, once the p value are null or extremely low, what shows clearly the disagreement between the judges.

For the analysis of the rotations we divided them in 3 sub-groups (RPIV - *relevé* rotations (pivot), RFF - rotations on the flat foot or on other part of the body and RF - *fouetté* rotations), because of their different characteristics that means different evaluation requirements (COP, 2012). In each sub-group of RPIV and RFF rotations, we analysed separately the basis of the rotation and the number of rotations associated to the basis.

The level of agreement of the judges evaluating the different type of rotations elements was, then, studied regarding the final ranking of the gymnasts (Table 7).

On the Table 7 we can see that for the basis of RPIV and RFF, there is no statistically significant difference between the evaluation done by the judges in the first and second parts of the ranking. We can see, also, that the values for significance drop substantially in the third part of the ranking. When we analyse the rotations associated to the basis of RPIV and RFF, we can see that in the first part of the ranking that the p value shows clearly the disagreement between the judges in evaluating such part of the difficulty.

Concerning the *fouetté* rotations, there is no agreement between the judges in the first and second parts of the ranking.

DISCUSSION

The goal of this study was determine the accuracy of the judges on the evaluation of each difficulty element presented in the difficulty forms.

Studying the forms in a global way we found that the percentage of elements where the 4 judges of panel agreed on the elements evaluation was higher than the disagreement cases. Nevertheless, we could observe that the judges agreed only in 60% of the elements, what is not enough for an evaluation that is supposed to be exact and accurate.

When we divided the gymnasts in 3 groups according to their place in the final ranking we found out that the judges showed a higher percentage of agreement for the gymnasts placed in the first part of the ranking and lower when we went down through the ranking. These results may suggest that it is more difficult for the judges to evaluate with precision the average and low level gymnasts. This evidence might be related to some criteria to validate the elements that, probably are not enough specific, what can cause some pliability in the evaluation. To solve this problem Simões (2000) suggests that all evaluation systems should hold precise criteria to allow judging correctly the performance. When the gymnast performs perfectly or almost perfectly the element, as usually happens with the top ranked gymnasts, is easier to the judges to recognize the difficulty, applying the evaluation criteria clearly, and tend to agree on its the evaluation. According to Bartolomeis (1999), the evaluation criteria are defined based on a successful criteria, which can facilitate the agreement of judges when the gymnast perform the elements with success, which is the case for the top ranked gymnasts. For the average and low level gymnasts is clearly more difficult to determine the "drop off" point to validate the difficulty elements because these gymnasts are doing the elements with some technical faults which leads the judges to struggle in applying the evaluation criteria stated in Code of Points (FIG, 2012).

We can also speculate that there could be an influence from what is expected, once the judges might expect better gymnasts to perform the difficulty elements correctly, as Findlay & Ste-Marie (2004) found, in a study with figure skating performances, that the judges gave higher scores to the better known skaters, comparing to the less known ones.

Other point that should be added to this discussion is the fact that the evaluation criteria for some difficulty elements include,

according to the Code of Points (FIG, 2012) some points concerning the quality of execution that may contribute to a higher variability on the validation of the elements. The interference of these execution quality criteria may create some variability in the work of the difficulty judge, creating some "grey zones" in the evaluation of difficulty elements. According to Askew (2002), the evaluator should direct all his attention for a specific profile and ignore the interference of any other information from a different profile.

The analysis of the results by apparatus revealed that the percentage agreement had not big differences for the routines performed with different apparatus. The results showed that behavior does not change from one apparatus to another; on the contrary we could remark that there was a consistency on the lake of accuracy in the difficulty elements evaluation. This consistency is due to the fact that the difficulty elements used in the different apparatus are basically the same and therefore, the requirements to validate the apparatus are the same (FIG, 2012).

Observing the results obtained for the judgment accuracy when we studied it for each apparatus and according to final ranking of the gymnasts we found out that the lower values of accuracy in the judgment were registered mainly in the gymnasts of the second part of the ranking. Besides what was already discussed about the lack of precision in defining the evaluation requirements, we are still able to speculate about the short amount of time that each judge has to consider a great amount of requirements defined for every single element in the routine composition, which may cause high variability between judges scores (Čuk & Karacsony, 2004). This is a problem for the average gymnasts because in opposite to higher level gymnasts where is easy to identify the difficulty elements correctly done and to lower level gymnast where visible when they do not perform the difficulty elements correctly, the average gymnasts often present an unclear version of the difficulty element making the decision to validate an element even more difficult than usual.

The results obtained when we analysed the level of agreement of the judges according to the type of difficulty element evaluated showed that the judges could not agree on the evaluation of the Mastery elements, and the Criteria (waves and pre-acrobatic elements) associated to the difficulty elements. These two groups showed levels of disagreement higher than the agreements, clearly in opposition to what happened with the other groups. The results suggests that definition of the evaluation requirements may have not an enough clear statement in the Code of Points (FIG, 2012), which can lead the judges in troubles to decide when the elements should be validate or not. According to the technical requirements to validate a Mastery element, it should be "a combination of extraordinary apparatus elements performed without technical faults". The definition of "extraordinary apparatus elements" is too vague to allow the judge to evaluate the elements with accuracy and could be also influenced by the international experience of the judge: after judging a certain number of international competitions the level of expectation for an "extraordinary element" can be raised. Knowing that in the World Championships the judges (one for each country participating) has different background experiences, we can understand that they cannot evaluate this technical requirement with same level of accuracy. In this way we strongly recommend that the Code of Points should include much more precise definitions of the technical requirements, because, according to Simões (2000) the evaluation criteria should be understood in equal manner by the various evaluators, in a way that the effect of the evaluation done may be valid and reliable.

After a more detailed analysis of each group of difficulty elements according to the gymnasts ranking, we could see that for the Jumps and Balances the level of agreement between judges was similar in the 3 parts of the ranking, showing that in these elements

the judges apply the same evaluation criteria. The evaluation criteria are understood and applied in the same way by the evaluators, once they produce the same result. This result allows us to speculate that visual image of the element allows a quicker and more reliable understanding, once the stated difficulties are presented. Boen, Karen, Yves, Jos, and Tim (2008) reach the conclusion that the possibility of feedback creates agreement between gymnastic judges. We know (unpublished study), that jumps and balances are repeated frequently in exercises, by the gymnasts in different apparatus routines, what facilitates the visual experience of the judge and therefore more precision in the application of evaluation criteria. According to Ste-Marie, Valiquette, & Taylor (2001), the visual image that is kept in the memory can influence the judge's performance. The agreement may be higher in the elements that appear often in exercises and because of that the judges have a clearer visual image and therefore a more precise evaluation.

In opposition, we can see that there are statistically significant differences between the 3 parts of the ranking in Mastery elements and DER elements, what clearly reveals the disagreement between the 4 judges on the validation of these difficulties. Besides what was already discussed above about the validation of Mastery elements, it is still relevant underline these elements are not listed and therefore the higher number of possible combination of handling contribute to make the evaluation of these type of elements even more difficult. We understand here that the absence of a list of Mastery elements would bring high improvements in routines creativity, although this could also bring the possibility for mixing originality concepts that should and must be evaluated in the originality item stated in COP (FIG, 2012). According to Balcells, Martín & Anguera (2009) it is possible to evaluate the originality and creativity with validity and reliability defining evaluation criteria that can be seen by the evaluators.

In the case of DER elements, the results lead us to the high number of criteria to bear in mind for the judge during the observation. According the Code of Points (FIG, 2012), the DER has an unlimited value and may contain till 19 different criteria that can be repeated. The judge has to memorize the criteria done to have the possibility to cross out on the difficulty form those what were not performed correctly or not done at all. Ste-Marie and Lee (1991) and Ste-Marie, Valiquette, & Taylor (2001) showed that the objectivity of a judge can be compromised by biases of memory. Also, the high number of criteria performed in such short may be responsible for this lack of agreement between the judges. We can speculate that the small amount of time that the judge has to observe and make all the possible deductions on the Difficulty form could be other source of variability between judges which may cause the evaluation of this group more vulnerable. Bucar, Čuk, Pajek, Kovac, & Leskosek (2013) and Čuk & Karacsony (2004) identified this same problem in the evaluation of the Vault execution in female artistic gymnastics, once this is also done in few seconds with 21 possible deductions.

The data concerning the Dance Steps showed also a significant disagreement of the judges in the validation. Dance Steps has, as criteria to be validate, the duration of at least 8 seconds, which can cause high variability in the evaluation, since this evaluation is done without a stopwatch or other device, but through the sensibility of the judge, and can be serious influenced by the *tempo* of the music.

The evaluation of the Mixed difficulties and Criteria associated to the difficulty elements (acrobatic elements and waves) reveals a significant disagreement between the judges, which could be due to the statement on the Code of Points concerning the link between the wave or acrobatic element and difficulty element itself. According to COP (FIG, 2012) the link must be immediately before or after but it is not clearly specified if it should be in continuity of the difficulty element or if it

could be a composition of two elements. According to Plessner (2005), the non-stated rules which can be considered as social norms, may influence the judge's decisions. It's important that they have great knowledge of the rules, to avoid wrong decisions.

Concerning the rotations, we can see that when evaluated the base of RPIV and RFF, there's no significant difference in the evaluation, in the first and second parts of the ranking. However, we can see that the values of a significant decrease in the third part of the ranking. Normally it is on the third part of the ranking where we find the lower level gymnasts and therefore with poor execution technique straight from the base of the rotation. According to the COP, the judge has to see the form, the degrees (360º) of the first turn and the technical faults that cancel the difficulty. The junction of all this factors (which are more present in the lower level gymnasts) belonging to two different profiles (difficulty and execution), may be explain the results of variability between judges found in the evaluation of this part of the difficulty.

Concerning the number of rotations associated to the base of RPIV and RFF, we can see that in the first part of the ranking there is clearly disagreement between the judges in evaluating these difficulty elements. About *fouetté* rotations, we found that in the first and second parts of the ranking there is no agreement between the four judges.

It is in the first and second parts of the ranking that the rotations performed done have a higher number of turns. By the evaluation criteria stated in COP, the judge has to count the number of full turns performed that is sustained fixed, without technical faults. Then, the difficulty in counting a high number of turns performed (that can go upper than 10 turns, mainly in *fouettés*) at high speed in few seconds, identifying the technical faults that implies the cancellation of the difficulty, may be in the origin of this variability for this kind of elements, in the first part of the ranking. Once again, we highlight here the interference of some criteria concerning execution, when judges are judging difficulty. According to Plessner (2005), positive and negative effects of prior knowledge on referee decisions and observation of a high amount of demand in such a short amount of time, may cause the loss of important information.

CONCLUSIONS

The four judges of difficulty panel did not agree in their evaluation in 40% of the difficulty elements presented in the difficulty forms. Regarding the final ranking of the gymnasts the agreement level is higher in the high and low level gymnasts. The level of accuracy was lower in the second part of the ranking, and in the difficulty elements which validation criteria depends not only from difficulty criteria but also from execution criteria.

The analysis by type of difficulty elements showed that for the Jumps and Balances the judges agreed on the evaluation of the elements which means an acceptable accuracy of judgement, but for the other types of elements the level of disagreement between the judges was significantly high to be an accurate judgement, where we highlight the Mastery and DER difficulty elements. This study provides updated information about the precision of difficulty judging in rhythmic gymnastics, to be considered in the possible alteration of the present code of points, in particular in the definition of the evaluation criteria of the elements where we see the highest disagreement between judges.

REFERENCES

Askew, S. (2002). Feedback for learning. *Journal of eduction for teaching*, *28*, 83-90.

Balcells, M., Martín, C., & Anguera, M. (2009). Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión Corporal y Danza Contact-Improvisatio. Apunts educación física y deportes.

Ciencias aplicadas a la actividad fisica y el deporte, 14-23.

Bartolomeis, F. (1999). *Avaliação e Orientação: Objectivos, instrumentos e métodos.* Lisboa: Livros Horizonte.

Bobo, M. (2002). *El juicio deportivo en Gimnasia Rítmica. Una propuesta de evaluación basada en indicadores de rendimiento. (PHD Thesis*), Universidad da Coruña, Instituto Nacional de Educación Física de Galicia, Coruña.

Boen, F. B., Karen, H., Yves, V. A., Jos, F., & Tim, S. (2008). Open feedback in gymnastic judging causes conformity bias based on informational influencing. *Journal of Sports Sciences*, *26*(6), 621–628.

Bucar, P. M., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? *Journal of Human Kinetics*, *37*, 173-181.

Čuk, I., Fink, H., & Leskošek, B. (2012). Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. *Science of Gymnastics Journal*, *4*, 73 – 82.

Čuk, I., & Karacsony. (2004). *Vault: methods, ideas, curiosities, history.* Ljubljana: STD. Sangvinck.

FIG. (2012). *Code of Points for Rhythmic Gymnastics Competitions.* Available at: http://www.fig-gymnastics.com/site/page/view?id=472

FIG. (2013). *Gymnastics Results*. Available at: http/www.gymnasticsresults.com

Findlay, C., & Ste-Marie, M. (2004). A Reputation Bias in Figure Skating Judging. *Journal of Sport & Exercise Psychology*, 26 (1), 154-166.

Gambarelli, G., Laquinta, G., & Piazza, M. (2012). Anti-collusion indices and averages for the evaluation of performances and judges. *Journal of Sports Sciences*, *30*(4), 411-417.

Heinen, T., Vinken, P., & Velentzas, K. (2012). Judging Performance In Gymnastics: A Matter Of Motor Or Visual Experience? . *Science of Gymnastics Journal, 4,* 63 – 72.

Kirkpatrick, J., & Hawk, L. (2006). *Curricula and evaluation: Maximizing results.* Measuring and Evaluating. Available from EBSCO .

Leandro, C. (2009). *Avaliação de Juízes de Ginástica Rítmica.* (Master Thesis), Porto University, Porto.

Leandro, C., Ávila-Carvalho, L., & Lebre, E. (2010). The avaluation of the performance of Rhythmic Gymnastics' Judges. Palestrica of the *Third Millennium Civilization & Sport, 11*(3), 202-206.

Lebre, E. (2011). *Technical principles for the new framework.* Crossroads to the Future [Press release].

Mark, M., & Shotland, R. (1987). *Multi Methods in Program Evaluation. New Directions For Program Evaluation* (Vol. 35). Londres: Jossey – Bass Inc.

Palomero, M. L. (1996). *Hacia una objetivación del Código*. Barcelona: Internacional de Gimnasia Ritmica Deportiva.

Plessner, H. (2005). Positive and negative effects of prior knowledge on referee decisions in sports. In T. Betsch, Haberstroh,S (Ed.) *The routines of decision making* (pp. 311–324).

Plessner, H., & Schallies, E. (2005). Judging the Cross on Rings: A Matter of Achieving Shape Constancy. *Applied Cognitive Psychology*, *19,* 1145-1145.

Simões, G. (2000). *A avaliação do desempenho Docente.* Lisboa: Texto Editora.

Ste-Marie, D., & Lee, T. D. (1991). Prior processing effect on gymnastic judging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 126-136.

Ste-Marie, D. M., Valiquette, S. M., & Taylor, G. (2001). Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. *Research Quarterly for Exercise and Sport*, *72*(4), 420-426.

Tamir, P. (1998). Assessment and Evaluation in Science Education . Opportunities to Learn and Outcomes *International Hand book of science*

Education (pp. 761-789): Dordrecht: Kluwer Academic Publishers.

Corresponding author:

Catarina Leandro
University Lusófona of Porto
Faculty of Psychology, Education and Sport
Rua Augusto Rosa, nº 24 (à Pç. da Batalha)
Porto 4000-098
Portugal
E-Mail: catarinaleandro@sapo.pt

Apendix 2. Published paper 2: The evaluation rules in the view of the rhythmic gymnastics judges.

Journal of Sports Science 4 (2016) 232-240
doi: 10.17265/2332-7839/2016.04.007

The Evaluation Rules in the View of the Rhythmic Gymnastics Judges

Catarina Leandro[1,3], Lurdes Ávila-Carvalho[2], Elena Sierra-Palmeiro[3] and Marta Bobo-Arce[3]

1. Faculty of Psychology, Education and Sport, University Lusófona of Porto, Porto 4000-098, Porto, Portugal

2. Faculty of Sports, University of Porto, Porto 4200-450, Porto, Portugal

3. Faculty of sport Science and Physical Education, University of Coruña, Coruña 15179, Oleiros, Spain

Abstract: The Code of Points (CoP) is the evaluation tool of Rhythmic Gymnastics (RG) and contributes to its evolution as sport. This tool is applied in competition by the judges who have a crucial role in the performance evaluation. The aim of this study was to characterize the International judges, their perception of the objectivity of their judgment in competitions and their opinions about the content and the reliability of the Code of Points. 162 international RG judges answered a questionnaire specially designed for this study. For the data analysis, non-parametric tests were used. According to the judges, the evaluation of the difficulty component in the individual routines has more subjectivity in the items of Mastery (58.6%) and Dance Steps (55.6%) and is less subjective in the body difficulties of Balance (72.2%). Concerning the execution, the judges consider that the evaluation of the artistic faults is the most subjective in this domain (64.8%). Within the artistic faults the items unity of the composition and relation between the Music and the Movements, were those which registered higher significant results for the subjectivity in the evaluation (47.5% and 37.0%, respectively).

Key words: Evaluation, objectivity, judge, rhythmic gymnastics.

1. Introduction

Performance the evaluation in RG depends on a judging process done by specialized judges that apply a specific set of rules and procedures established in the official FIG Code of Poins (CoP) [1]. The CoP is used as an evaluation instrument by the judges, who have their own technical, human and social experiences. All these aspects take part in the judging process and from it depends, in part, the improvement of quality of the sports practice, the safeness of the physical and moral integrity of the athletes and the reenforcement of ethical values in sport.

Thus, the judge's background is a factor of sportive quality. The performance of international level judges overcomes the performance of less experienced judges, once they use of other cognitive strategies, increasing their global efficiency of error spotting [2]. Also, "If we seek sports excellence, it is not possible not to understand the job of judgement and refereeing in sport. It is not enough to train coaches with scholar degrees, masters and PhDs, it's necessary to think with vision, because it is definitely incongruent and even not logic for the one who evaluates the process and final result of the coach work, not to hold an integral formation that amounts to the level in which the process is occurring" [3]. Academical, professional and social formation of the referee is a characteristic acknowledged by them as being a guarantee of their performances quality [4].

On the other side, the CoP which rules and orientates all the judges' actions, works as an evaluation tool that depending on its structure, its content, and its reliability. All this factors may have a better or worse impact on the judges' performance as evaluators. It is concluded that the most valued skills are those related to the sport's technical parameters and the ability to adapt to any level of competition with self-confidence and self-assuredness [5]. Considering that the judge and the

Corresponding author: Catarina Paula Leandro Sousa Silva, Ph.D., professor, research fields: evaluation in sport.

CoP hold an inseparable dialectic, it is necessary to analyze them together. So the aims of this study are: to define the population of the "International RG Judges" according to personal information, education, professional experience and experience as judges; and to identify their opinion about the CoP 2012 related to its content, structure, clarity and validity of the rules to be applied, as well as possible changes to be introduced to contribute to and improvement of the judging and therefore the correct evolution of the sport.

2. Methods

162 international RG judges answered a specific questionnaire specially developed for this study. It was composed by 15 questions grouped in 2 categories: (1) personal information, education, professional experience and experience as a judge, and (2) objectivity of evaluation in RG and proposals to change the Code of Points [1].

This study was approved by the International Gymnastics Federation (FIG). All the international RG judges, from all over the world, 287 judges were invited by FIG to answer the questionnaire available at Google Drive.

To protect the judges' anonymity, the answers were received anonymously on google drive, so the full blinding of the judges involved was undertaken. The data was collected between July and September 2014.

2.1 Statistical Analysis

For the statistical analysis, we used the Statistical Package for the Social Sciences-Version 21.0 (SPSS 21.0, Chicago, USA) and Microsoft Office Excel 2010.

For the data analysis, non-parametric tests were used (Friedman test and Sign test between groups) to determine whether there were significant differences between groups. Significance level was set at $\alpha = 0.05$ (corresponding to a confidence level of 95%). The frequencies and percentages of the prognostic variables were calculated through the descriptive statistics.

3. Results

The characterization of the Judges is resumed in Table 1.

Table 1 Descriptive statistic of the judges characterization.

Characterization of Judges (N = 162)					
				Freq.	%
Personal information	Sex	Female		160	98.8
		Male		2	1.2
	Age	Mean	43.4		
		Minimum	22		
		Maximum	68		
	Country			59	
Education	High School			16	9.9
	University			80	49.3
	Master			51	31.4
	PHD			15	9.4
Work Experience (RG Coach)	Yes			141	87
	No			21	13
RG International Judge	Brevet I			6	3.7
	Brevet II			29	17.9
	Brevet III			66	40.7
	Brevet IV			61	37.7
RG International Judge Experience	Less than 1 Olympic cycles			42	25.9
	1-2 Olympic cycles			51	31.5
	More than 2 Olympic cycles			69	42.6

The judges were 43.4 years old, 49.3% have a university degree, 87% are also coaches, 40.7% are judges brevet III and 42.6% have been judges for more than 2 Olympic cycles.

We can see in Table 2 the summary of the collected data about the judges' opinion about the objectivity of Difficulty evaluation. The *Mastery* is considered to be the item with less objectivity in evaluation with the answer "less objective to evaluate" collecting 58.6% of the answers. Following we have, in increasing order for the objectivity in judgement, the following groups: *Dance Steps*, *DER (base)* and *Rotation*s in which the answer "more or less objective to evaluate" is the more frequent one 55.6%, 45.7% and 51.2%, respectively. In the groups *DER (criteria)* and *Jumps* the answer "objective to evaluate" is the more frequent with 40.7% and 46.3% respectively. Yet, it's in *Rotations (Basis)* and *Balances* that we see higher values of objectivity in the evaluation with the answer "objective to evaluate" getting 59.9% and 72.2% of the answers, respectively.

Fig. 1 presented the average indicator of objectivity in the evaluation of the different difficulty groups according to the judges' opinion (groups were written in increasing order within the indicator). Globally, we can state that there's a statistically relevant difference (Friedman test, $P = 0.000$) in the objectivity for the different difficulty groups.

Comparing groups, we can see in Table 3 that the *Mastery* show a significant difference ($P = 0.000$ from all of the others groups, being the Mastery the group where the evaluation is seen by the judges as less objective. Also, between *Rotations (base)* and *Jumps*, there are statistically significant differences ($P = 0.002$), being the evaluation less objective in *Jumps* rather than in *Rotations (base)*. The same happens between the groups of *Balance* and *Rotations (base)* ($P = 0.010$), where the evaluation is less objective in *Rotations (base)* than in *Balance*.

When comparing in pairs the groups *DER (base)*, *Rotation (Rot Add)*, *DER (Criteria)* and *Jumps* we can see that there are no statistically significant differences on the degree of objectivity in evaluation. We can even say that the objectivity in evaluation as seen by the judges holds a similar distribution in the four groups ($P = 0.117$).

We can see in Table 4, the summary of the collected data from the judges' opinion about the objectivity in evaluating Execution in Technical and Artistic Faults.

The Technical Faults item is considered the one with higher objectivity in evaluation with the answer "objective to evaluate" getting 80.2% of the answers and the Artistic Faults item is considered the one with less objectivity in evaluation with the answer "Less objective to evaluate" holding 64.8% of the answers. The difference in objectivity in the evaluation of Artistic Faults and Technical Faults is statistically significant (Friedman test, $P = 0.000$).

Regarding the Artistic Faults sub-items (Table 5), the Unity of Composition item is considered the less objective one in evaluation with the answer "less objective to evaluate" collecting 47.5% of the answers. Next, in ascending order of objectivity, the items Music/Movement, and Body Expression were considered "more or less objective to evaluate" (45.7% and 56.6%, respectively). The higher values of objectivity in evaluation are seen in the item Use of Space with the answer "Objective to evaluate" getting 54.3%. Globally, there are statistically significant differences (Friedman, $P = 0.00$) in the objectivity of the evaluation of the different items within the Artistic Faults group, being the degree of objectivity higher in some items than others.

The Artistic Faults item is considered the one with less objectivity in evaluation with the answer "Less objective to evaluate" holding 64.8% of the answers. Regarding the Artistic Faults sub-items, the Unity of Composition item is considered the less objective one in evaluation with the answer "less objective to evaluate" collecting 47.5% of the answers. Next, in ascending order of objectivity, the items Music/Movement, and Body Expression were considered

Table 2 Descriptive statistics of the judges' opinion about the objectivity in difficulty evaluation.

Objectivity of Evaluation	Mastery	Dance Step	DER (Base)	DER (Criteria)	Jumps	Balance	Rotation (Base)	Rotation (Added)
	Frequency Tables (%)							
Less objective	58.6	22.8	21.0	21.6	11.1	3.1	7.4	16.0
More or less objective	34.0	55.6	45.7	37.7	42.6	24.7	32.7	51.2
Objective	7.4	21.6	33.3	40.7	46.3	72.2	59.9	32.7

Friedman test $P = 0.000^{*}$.

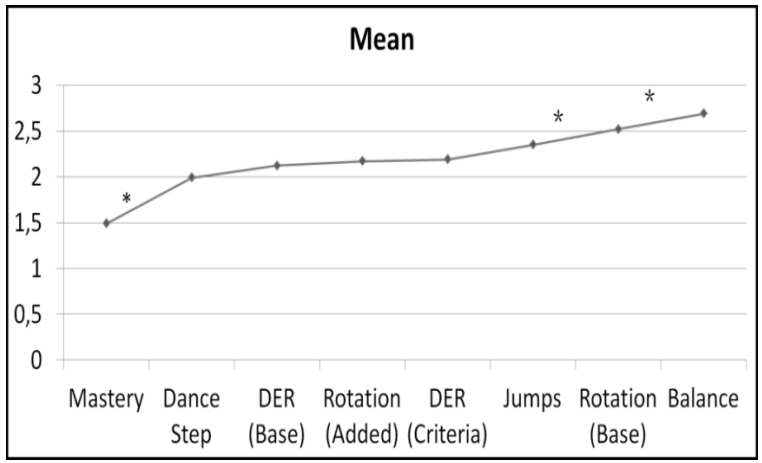


Fig. 1 Average indicator of objectivity in evaluation in the different difficulty groups.

Friedman test $P = 0.000^{*}$.

Table 3 Sign test between difficulty groups by the objectivity in the evaluation.

Sign Test							
	Dance Step-Mastery	DER (base)-Dance Step	Rot (Added)-DER (Base)	DER (crit.)-Rot. (Added)	Jump-DER (crit.)	Rot. (Base)-Jump	Balance-Rot. (Base)
Z	-6.904	-1.724	-0.625	-0.256	-0.891	-3.086	-2.585
Asymp.sig	0.000*	0.085	0.532	0.798	0.373	0.002*	0.010*

"more or less objective to evaluate" (45.7% and 56.6%, respectively). The higher values of objectivity in evaluation are seen in the item Use of Space with the answer "Objective to evaluate" getting 54.3%. Globally, there are statistically significant differences (Friedman, $P = 0.00$) in the objectivity of the evaluation of the different items within the Artistic Faults group, being the degree of objectivity higher in some items than others.

According to the judges' opinion, we can see in Fig. 2 the average indicator of objectivity in the evaluation of Artistic and Technical Faults (groups are in ascending order within the indicator).

When we try to analyze if there are significant differences between the items distributions, we see in Table 5 that the item Unity Composition differs from the item Music Movement ($P = 0.002$), with also a significant difference from all the others items, once it's in the item Unity Composition that the evaluation is seen by the judges as less objective. Also between the items Body Expression and Music Movement, there are statistically significant differences ($P = 0.000$), being the evaluation less objective in Music Movement than in Body Expression. The same happens between the items Use of Space and Body Expression ($P = 0.017$), where the evaluation is less objective for Body Expression than for Use of Space.

Concerning the Difficulty, in the opinion of the judges, the evaluation criteria for Mastery should be changed, holding 69.8% of the answers (Fig. 3). The

Table 4 Descriptive statistics of the Execution faults by the objectivity in the evaluation.

Execution Faults		
	Technic Faults	Artistic Faults
Objectivity of Evaluation:	*Frequency (%)*	
Less objective	1.9	64.8
More or less objective	17.9	29.6
Objective	80.2	5.6

Friedman Test $P = 0.000$.

Tabela 5 Descriptive statistics of the Artistic Faults by the objectivity in the evaluation.

Artistic Faults				
	Unity Composition	Music/Movement	Body Expression	Use Space
Objectivity of Evaluation:	*Frequency (%)*			
Less objective	47.5	37.0	16.7	23.5
More or less objective	42.0	45.7	56.8	22.2
Objective	10.5	17.3	26.5	54.3

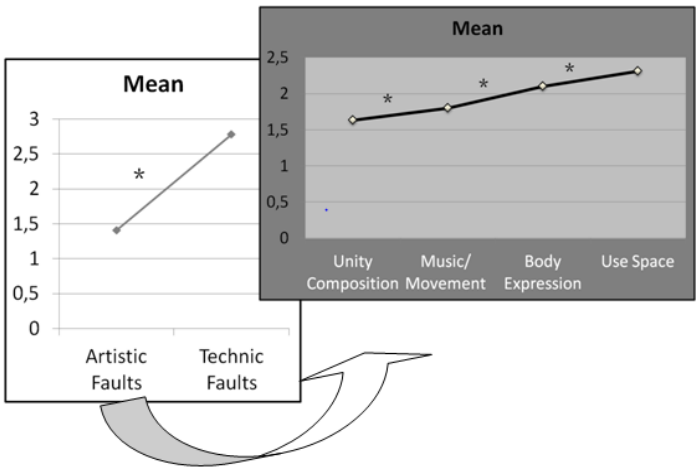


Fig. 2 Average indicator of objectivity in the evaluation of Artistic and Technical Faults.
Friedman test $P = 0.000^{*}$.

Table 5 Sign test between Artistic Components by the objectivity in the evaluation.

	Sign Test		
	Music Movement-Unity Composition	Body Expression-Music Movement	Use Space-Body Expression
Z	-3.125	-4.533	-2.388
Asymp.sig	0.002*	0.000*	0.017*

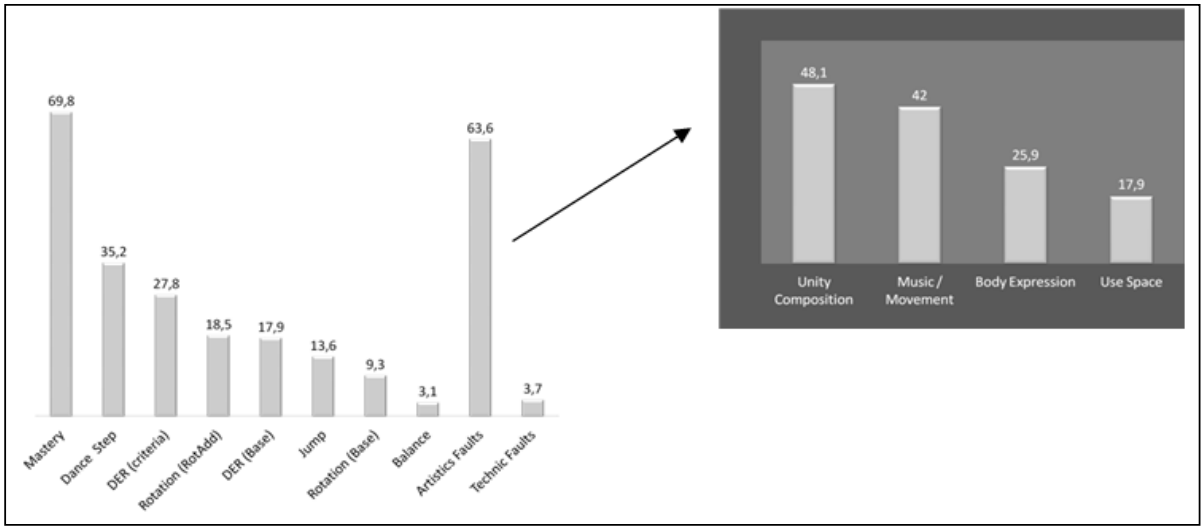


Fig. 3 Frequency table of the difficulty and execution groups by the judges opinion be changed.

Table 6 Frequency of the judges' opinion ("agree that there should be a limit of repetitions for the body difficulty in the different apparatus").

Frequency Table		
	Frequency	%
NOT	57	35.2
YES	105	64.8
Total	162	100.0

Balances group is the one that should not suffer any alteration in the evaluation criteria, holding 3.1% of the answers.

For the Execution, the Artistic Faults evaluation criteria should be changed (63.6%), in opposition to the Technical Faults, which got only 3.7% of the answers.

Studying the items of the Artistic Faults, the evaluation criteria for the Unity of Composition and Music/Movement should be changed, holding 48.1% and 42% of the answers, respectively. The item Use of Space is the one that should not suffer any alteration in the evaluation criteria, holding 17.9% of the answers.

We can see in Table 6 that the majority of the judges (64.8%) agree that there should be a limit of repetitions for the body difficulty in the different apparatus, to improve the variety and the composition of RG routines.

4. Discussion

The sportive judgement, as human, individual and collective action, holds intellectual, volitional and ethical components, which should be taken into consideration in an integrated global way, so the evaluation of sport performance is done in a responsible manner [6].

In this study, we can verify that the majority of international judges has a high academic level, works as a RG coach and has a large experience in judgement. This type of background offer efficiency conditions to the judges, since they register quality judging values, when compared to quality values of judgement got by judges with less experience, as well as lower academic level [7].

The criteria that can distinguish a "specialist in the matter" goes from the connection of the specialist with

the problem, the professional experience, the personal qualities or professional ability, to the guaranty of the quality of the answers and the skill of recognizing detailed information [8].

The most significative difficulties in judgement, come from the excessive amounts of information that the judge needs to quickly summarize in order to transform into a score; in a practical level the judges are able to solve their problems thanks to their experience and knowledge about gymnastics [9], which also allows us to think that the results support the idea that RG judgment is done by individuals who, besides their knowledge of the scoring code, have other resources such as experience and insight, which could be a *plus* in the judgement once the experiences and global vision of the sport lead the judges to deduce, from logic, some aspects that maybe hard to identify or differentiate by younger or inexperienced judges.

Besides de characterization of the judges, we tried to know what they think about the CoP, once they are the ones who use it as an evaluation tool in RG routines. When considered the indicator "objectivity in evaluation" in the different difficulty groups under analysis, the results showed that globally there are statistically significant differences in the objectivity of the evaluation, being the degree of objectivity superior in some groups compared to others. The group with less objectivity in evaluation is Mastery. Following in increasing order of objectivity are Dance Steps, DER (Basis), Rotations (Additional rotations) and DER (criteria).

These results may suggest that the judges find more difficult to evaluate with precision some elements performed by the gymnasts, since their opinion about the evaluation of these groups is less objective probably due to the way the evaluation criteria is described in the CoP, allowing different interpretations.

The evaluation of human performance for some sports is not possible to be done through mechanical ways [10], thus making the reference of the pattern criteria the way to assure validity and reliability in the result of the evaluation, when trying to evaluate the quality of a movement.

It is also important to state that the complexity in the evaluation of the referred difficulty groups (Mastery, Dance Steps and DER) may be also related to the fact that there is no pattern reference in the CoP, compared to what happens for other difficulty groups such as Jumps, Balances and Rotations.

The difficulty groups in which the evaluation is more objective are Jumps, Rotations (Basis) and Balances. These results indicate that in the opinion of the judges, the evaluation criteria of these groups described in CoP allow an objective evaluation, with easy application. It is important to state that the CoP holds a list with pattern images for these three groups, which in our opinion allow an immediate perception of the correctly performed difficulty and consequently the objectivity in evaluation. The criteria to determine the quality of the evaluation should refer to a pattern, model or arbitrary level of that same quality [11].

The study about the degree of agreement between the 4 judges in the evaluation of the different difficulty groups declared in the competition cards used in KIEV 2013 WC, supports the results obtained here [12], once this study highlights the same difficulty groups, where we see more disagreement in the evaluation done by the judges.

When considering the indicator "objectivity in evaluation" within the Execution items, the results showed that globally that there are significant differences in the evaluation objectivity of Technical Faults and Artistic Faults. The Artistic Faults group is considered the less objective in the evaluation. About this group, we found out that there are statistically significant differences in the objectivity of evaluation of the different items that integrate it, with higher objectivity for some items rather than others. The parameter Unity of Composition is considered the less objective in the evaluation. Following in increasing order of objectivity there are Music/Movement and

Body Expression. It's the Use of Space item where we get higher values of objectivity. In the same way, Bučar et al. [13] found results of low validity and reliability in the judgment of artistic components in Artistic Gymnastics, what allows us to consider that the current instruments of evaluation for gymnastics artistic components require monitoring for a possible reassesment and eventual restructuring.

We also tried to identify which evaluation criteria would the judges like to modify in order to potentiate their performances as evaluators. The results found show proposals of changes in the evaluation criteria for Mastery group and for the Artistic Faults group, in particular for "Unity Composition" and "Music/Movement" parameters. Finally, the results indicate that the majority of the judges (64.8%), consider that the CoP should limit the repetitions of body difficulties in the different apparatus, helping to enrich the compositions of RG routines and consequently the evolution of the sport. Liu et al. [14] analyzed the evolution of scoring codes in RG and found out that the evolutive tendency should contemplate variety and diversity allowing exploration of new skills.

5. Conclusions

The evaluation system to determine the final scores of a RG exercise is given by the CoP, being this an instrument elaborated by the FIG. Yet, the judges are the ones using it as an evaluation tool thus their opinions represent a reference value to be considered in its elaboration. They manifested different opinions about the objectivity in the evaluation of the Difficulty, Execution, as well as the different parameters of evaluation in artistic faults.

They highlighted the Mastery, Dance Steps and DER, in Difficulty and the Artistic Faults (mainly "Unity Composition" and "Music/Movement") in execution, as being the ones with most complexity in evaluation when considered the objectivity. They suggested changes in the evaluation criteria of these groups, in order to become more precise in the final evaluation.

Finally the judges stated that they would like to have in the CoP some rules to limit the repetition of difficulties in the different apparatus routines, in order to improve the diversity and variety in RG routines, promoting an enrichment of the sport.

Therefore, we conclude that the instrument of evaluation used right now is not yet ideal to absolutely assure the validity and reliability in RG judgment. These results may contribute for a reconstruction of the CoP and consequently help in the evolution of the sport.

We expect that new rules of artistry evaluation will bring improvement of reliability and consistency of judges and this should be verified through further research of future competitions.

Acknowledgments

We are especially grateful to International Gymnastics Federation (FIG). We also thank to international Rhythmic Gymnastics Judge.

References

[1] FIG-International Gymnastics Federation. 2012. "Code of Points for Rhythmic Gymnastics Competitions, 2012-2016." Accessed November 18, 2012. http://www.fig-gymnastics.com/site/page/view?id=472

[2] Flessas, K., and Mylonas, G. 2015. "Judging the Judges' Performance in Rhythmic Gymnastics." *Medicine & Science in Sports & Exercise*. *American College of Sports Medicine* 47 (3): 640-8.

[3] Guardo, M. 2004. "Toward a Theory of Arbitration for Sport." *Rev*. *Digital*. *Buenos Aires 68*. Accessed July 8, 2004. http:/www.efdeportes.com/ (in Spanish)

[4] Martin, S. 2006. "Desirable Characteristics for Arbitration in Sports Trial in Judo". *Latin American Journal on Exercise and Sport Psychology* 1: 27-40. (in Spanish)

[5] Fernandez-Villarino, M. 2013. "Practical Skills of Rhythmic Gymnastics Judge." *Journal of Human Kinetics* 39: 243-249.

[6] Palomero, M. L. 1996. "Towards an Objectification of the International Code of Points Rhythmic Gymnastics." Ph.D. thesis, The Barcelona University. (in Spanish).

[7] Leandro, C., Ávila-Carvalho, L., and Lebre, E. 2010. "The Evaluation of the Performance of Rhythmic Gymnastics'

Judges." *Palestrica of the Third Millennium Civilization & Sport* 11 (3): 202-6.

[8] Almenara, J. 2013. "The Use of Expert Judgment for the Evaluation of TIC: The Coefficient of Expert Competence." Ph.D. thesis, Sevilha University. (in Spanish).

[9] Plessner, H. 2005. "Positive and Negative Effects of Prior knowledge on Referee Decisions in Sports." In *The Routines of Decision Making*, edited by Betsch, T., and Haberstroh, S. 311-24.

[10] Morrow, J., Jackson, A., G-Disch, J., and Mood, D. 1995. *Measurement and Evaluation in Human Performance*. Library of Congress Mystat, USA.

[11] Simões, G. 2000. *The Evaluation of Teacher Performance*. Texto Editora, Lisboa, Portugal. (in Portuguese).

[12] Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., and Bobo-Arce, M. 2015. "Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines." *Science of Gymnastics Journal* 7 (3): 81-93.

[13] Bučar Pajek, M., Kovač, M., Pajek, J., and Leskošek, B. 2014. "The Judging of Artistry Components in Female Gymnastics: A Cause for Concern?" *Journal of Human Kinetics* 37: 173-81.

[14] Liu, X. X., and Kuang, L. 2001. "Review of Evolvement Course of International Evaluation Rules in Rhythmic Gymnastics and Its Effects on Technique Development." *Journal of Beijing Sport University* 3 (24): 412-5.

Apendix 3. Published paper 3: Departure difficulty score vs final difficulty score. The effect of performance in elite rhythmic gymnastics.

Athens Journal of Sports *X Y*

Departure Difficulty Score Vs Final Difficulty Score. The Effect of Performance in Elite Rhythmic Gymnastics

By Catarina Leandro[*]
Lurdes Ávila-Carvalho[†]
Elena Sierra-Palmeiro[‡]
Marta Bobo-Arce[+]

The aim of this study was to determine how the difference between the departure difficulty score, and the final difficulty score, in the different type of elements affects the performance success in elite rhythmic gymnastics. Two hundred eighty-eight individual routines performed at the World Championship 2013 were analysed. A non-parametric test- Kruskal-Wallis test, was applied to determine whether there were significant differences between different level groups. The results show significant differences between the departure difficulty score and final difficulty score, which increase as the level of performance decreases. Rotations, DER and Mastery are the main types of difficulty elements responsible for this difference. These results suggest that the judges and coaches do not have the same perception of the evaluation criteria of the difficulty elements. The findings can contribute to improve the definition criteria of the difficulty elements and to clarify the specific needs of the training program.

Keywords: *Difficulty score, coach, judge, rhythmic gymnastics.*

Introduction

The ability to control and monitor the technical content of competitive individual routines in Rhythmic Gymnastics (RG) accurately is an important factor of effective high performance in this sport. This knowledge can also contribute to preview and characterise the effort requirements allowing improvements in a gymnasts' preparation for elite competitions (Ferreirinha, 2011). A precise understanding of the technical content of individual RG routines, acknowledged by the judges' evaluation, can be beneficial to the International Gymnastics Federation (Cuk et al., 2012), as well as for coaches and gymnasts that can also be used as feedback to adapt the training structure in order to improve performance (Fernandez-Villarino, 2013).

Several studies (Caburrasi, 2003; Ávila-Carvalho, 2011; Ávila-Carvalho, 2012; Trifunov, 2013; Agopyan, 2014) analysed the number and the level of

[*] Faculty of Psychology, Education and Sport, University Lusófona of Porto, Portugal and Faculty of sport Science and Physical Education, University of Coruña, Spain.
[†] Sports Faculty, CIFI2D, University of Porto, Portugal.
[‡] Faculty of sport Science and Physical Education, University of Coruña, Spain.
[+] Faculty of sport Science and Physical Education, University of Coruña, Spain.

difficulty elements presented by the coach on the competition form. The number and the level of difficulty elements is the departure difficulty score (DDS). However, those studies did not emphasise information about the contribution of each type of element acknowledged by the judges, which leads to the final difficulty score (FDS).

The elements prescribed before the competition in the forms do not translate the real success in competition, which suggests that the use of these indicators may not be enough to understand the individual adaptations and to establish the optimised training models (Arkaev, 2004). Also as a consequence of the constant and quick evolution of this sport, a permanent upgrade of these kinds of studies is needed. (Caburrasi, 2003; Cuk, 2012; Massidda, 2012; Hokelmann et al., 2012; Bucar, 2013, Pelin, 2013).

An individual RG routine is composed of body and apparatus elements linked in a specific way, which are called difficulty elements. The code of points (CoP) holds a great variety of difficulty elements to be used in the routines. One important characteristic of the RG is to allow the gymnast to link it in her own way, with a clever configuration, and a stylish and prefect presentation (Wang, 2013). On the present Olympic cycle, the content of an individual RG routine should respect the specific requirements that are common to the routines of all 4 apparatus: Jumps/Leaps, Balances, Rotations, Apparatus Mastery, Dance Steps and Dynamic Elements with Rotation and throw (DER) (FIG, 2012). The value of each difficulty element is from 0.10 points to 1.50 points or more, which may be absolutely determinant in the final score obtained in the competition. The inclusion of complex abilities in the routines is essential to have a high score (Massida, 2012).

It is important to understand which groups of elements contribute the most for the difference between the DDS proposed by the coach on the difficulty form, and the FDS given by the judge in competition. Once the coaches know the cause of this difference, they may optimise the training process on these groups of elements, promoting an effective success in competition. Thus, before suggesting eventual future changes, it is important to find out what should be changed and what should be maintained and how.

The aim of this study is to determine the real performance success in elite rhythmic gymnastics through the analysis of the difference between the departure difficulty score and the final difficulty score. The analysis will be done (i) according to the final ranking of the gymnast in the competition in order to see if the technical level of the gymnasts has influence in the results, and (ii) according to the type of element, to determine if there are elements with more influence in the difference between the DDS and the FDS.

Material and Methods

Participants

One thousand and one hundred and fifty-two difficulty forms concerning 288 individual routines were analysed (4 forms per routine, 1 per judge). The routines were performed by gymnasts from 45 different countries competing at the Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013. The final scores were obtained from the official book of results of the qualification competition (FIG, 2013).

Measures/ Procedure

All difficulty elements reported in the difficulty forms provided by the gymnasts at the competition were recorded. The evaluation of the each difficulty element was considered according to the average of the 2 intermediate scores done by the 4 judges on the form. The analysis was done considering the sample clustered into 3 groups according to the gymnasts' final ranking as follows: The top 24 gymnasts on the ranking (Group 1), the 24 middle gymnasts on the ranking (Group 2) and the 24 lower placed gymnasts on the ranking (Group 3). This division allowed the comparison of the routine difficulty value declared by the coach in the difficulty form (DDS) with the difficulty score given by the judges during the competition (FDS), for gymnasts with different technical levels. Then, we studied the sample according to the type of difficulty elements performed, listed according to the composition requirements of the Code of Points (FIG, 2012).

Full blinding of the judges and gymnasts involved was undertaken. That is, in order to protect the judges' and gymnasts' anonymity we blinded their names.

The forms were analysed by two international RG judges. The intraclass correlation coefficient (ICC) in the test-retest method (intra-examiner) was 0.99. The ICC between the observers (inter-examiner) was 0.98.

Statistical Analysis

The data were analysed using the Statistical Package for Social Sciences – version 20.0 (SPSS 20.0, Chicago, USA) and Microsoft Office Excel 2007. The level of significance was set at $\alpha = 0.05$ (confidence interval of 95%). Descriptive statistics were calculated using the mean values as a measure of central tendency, standard deviation (SD) as a measure of dispersion. After checking the normalities in the data distribution ($p<0.05$) using the Kolmogorov-Smirnov normality test, we resorted to a non-parametric test-Kruskal-Wallis test, to determine whether there were significant differences between the three groups in the Rhythmic Gymnastics World Championship ranking. A multiple regression was used to analyse the influence of each difficulty element in the gymnasts' final difficulty score.

Results

Comparing the routine difficulty value declared by the coach in the difficulty form (DDS) with the difficulty score given by the judges during the competition (FDS) we get the results summarized in Table 1 where we present the average values for DDS and FDS (mean ± sd) and the difference (Δ) between these values in the 3 groups of the gymnasts' final ranking.

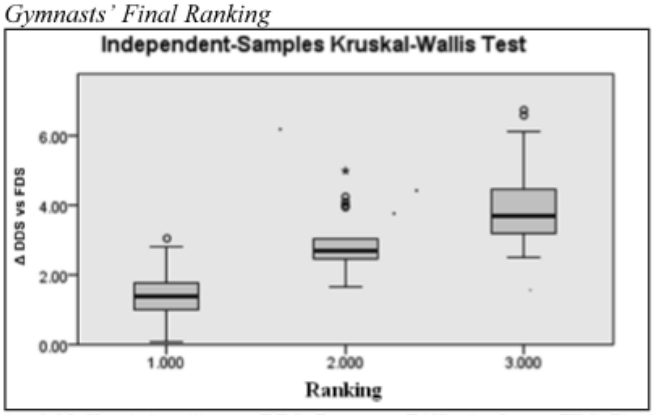
Table 1. *Descriptive Values of Departure Difficulty Score and Final Difficulty Score*

Routine Difficulty Score	**Group 1 (n=96)**	**Group 2 (n=96)**	**Group 3 (n=96)**
	Mean±sd	**Mean±sd**	**Mean±sd**
DDS	9,83±0,22	9,59±0,26	8,50±0,98
FDS	8,42±0,64	6,81±0,62	4,61±1,37
Δ DDS vs FDS	1,41±0,58*	2,78±0,57*	3,88±0,94*

*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score; Δ DDS vs FDS: differences between DDS and FDS
Source: Authors' estimations.

We could see, by only observing the numbers, that there is a considerable difference between the DDS and the FDS, even in the 24 best gymnasts (1.41±0.58 points). This difference is almost twice as high in group 2, the 24 middle gymnasts (2.78±0.57 points) and for group 3, it achieves an average of 3.88±0.94 points. We found statistically significant differences between these results, visible in Figure 1.

Figure 1. *Boxplot for Difference between DDS and FDS for the 3 Parts of the Gymnasts' Final Ranking*



*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score; Δ DDS vs FDS: differences between DDS and FDS

After getting these results we analysed the differences between the DDS and FDS, considering the different types of difficulty elements, trying to find out if the differences are mainly connected to some type of difficulty elements, and if the type of elements with a higher difference are the same for the 3 groups of the gymnasts ranking. In Table 2 we can observe the DDS and FDS for each type of difficulty element for the gymnast in the different groups.

Table 2. *Descriptive values of the different types of the difficulty elements of the Departure Difficulty Score and Final Difficulty Score*

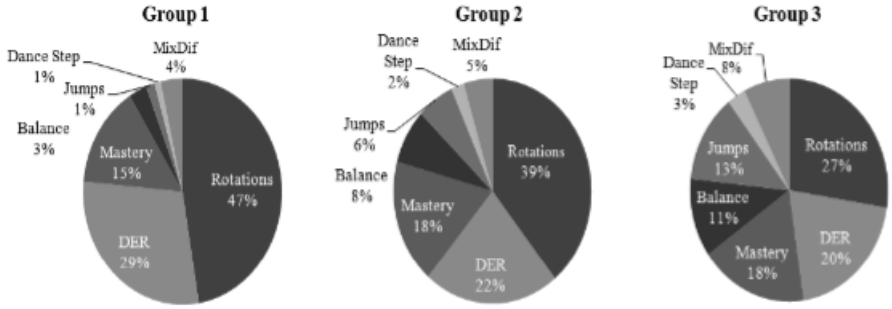
Elements Difficulty Score		Group 1 Mean±sd	Group 2 Mean±sd	Group 3 Mean±sd
Jumps	DDS	1,54±0,44	1,40±0,45	1,25±0,33
	FDS	1,52±0,44	1,23±0,48	0,72±0,49
	Δ DDS vs FDS	0,01±0,07*	0,16±0,25*	0,52±0,43*
Balance	DDS	1,09±0,51	1,18±0,45	1,16±0,40
	FDS	1,05±0,52	0,96±0,46	0,69±0,38
	Δ DDS vs FDS	0,04±0,11*	0,21±0,28*	0,47±0,40*
Rotations	DDS	3,00±0,88	2,89±0,66	2,22±0,71
	FDS	2,32±0,77	1,80±0,47	1,08±0,52
	Δ DDS vs FDS	0,67±0,37*	1,08±0,42*	1,13±0,48*
DER	DDS	2,05±0,30	2,01±0,27	1,83±0,35
	FDS	1,64±0,42	1,40±0,38	1,00±0,51
	Δ DDS vs FDS	0,41±0,37*	0,61±0,33*	0,83±0,44*
Mastery	DDS	0,77±0,40	0,82±0,45	0,96±0,71
	FDS	0,56±0,40	0,32±0,27	0,24±0,27
	Δ DDS vs FDS	0,20±0,18*	0,50±0,31*	0,72±0,59*
MixDif	DDS	0,98±0,20	0,90±0,10	0,88±0,05
	FDS	0,93±0,27	0,76±0,25	0,57±0,28
	Δ DDS vs FDS	0,05±0,13*	0,13±0,24*	0,31±0,30*
Dance Steps	DDS	0,81±0,26	0,76±0,23	0,81±0,30
	FDS	0,79±0,26	0,70±0,23	0,69±0,30
	Δ DDS vs FDS	0,01±0,07*	0,05±0,12*	0,12±0,18*

*p<0.05 Kruskal-wallis test; DDS: Departure Difficulty Score; FDS: Final Difficulty score; Δ DDS vs FDS: differences between DDS and FDS
Source: Authors' estimations.

As we can see in Table 2, the difference between the DDS and FDS is highly correlated with the final ranking of the gymnasts in each difficulty element. The lowest is the position of the gymnast in the ranking and the highest is the difference between the departure and the final score in all types of difficulty elements. We can distinctively see the difference between the DDS and FDS behaviours according to the different types of difficulty elements. For the Jumps, Balances, MixDif and Dance Steps the difference between the DDS and the FDS is very low for the best gymnasts and increases a lot as we go down in the gymnasts' ranking. For the Rotations, DER and Mastery the difference between the DDS and the FDS is very high in the 3 groups.

Figure 3 shows the contribution of each type of element (in percentage) to the difference between the DDS and the FDS according to the final ranking of the gymnast.

Figure 3. *Contribution of Each Type of Element (in percentage) to the Difference between the DDS and the FDS, for the 3 Parts of the Gymnasts' Final Ranking*



We can see that the Rotations clearly contribute, in the 3 groups of the ranking (47%, 39% and 27%) to the difference between DDS and FDS. It is worth highlighting that the Rotations together with the DER group are responsible for more than half of the difference between the DDS and FDS for the gymnasts in group 1 and group 2 of the ranking (76% and 61% respectively). The Dance Steps is the element that clearly shows smaller differences between the DDS and FDS for the gymnasts placed in the 3 groups of the ranking.

Discussion

According to the results we can see that there are large differences between the scores proposed by the coaches in the competition forms and the difficulties that judges could identify, which increase as the gymnasts go lower in the ranking, going from almost more than 1 point in the gymnasts placed in group 1 of the ranking, to almost 4 points for the gymnasts placed in group 3 of the ranking, with a statistically significant difference between the 3 levels of gymnasts. Analysing these results we can remark that if we could expect differences between the DDS and FDS in the weaker gymnasts due to the faults in execution which cancel the value of the difficulty (FIG, 2012), the same result would not be expected in the very good gymnasts (the first 24 in the ranking), since these gymnasts get higher scores in execution (Leviotti, 2012).

The average of the DDS presented in the competition was 9.3 points, close to the maximum grade of 10 points. This average suggests a high global stage of world excellence, but truthfully, as we can see in Table 1, not even the best-ranked gymnasts in the group one of the ranking can reach this score.

Therefore, we can state that the difficulty value of the proposed routine by the coach in the competition form (DDS) is ambitious and does not reflect the performance capacity of the gymnast.

These results could be seen in two ways: on the one hand, perhaps all coaches "overwrite" the competition forms knowing that the judges "feel" the need to cut some difficulties (Ávila- Carvalho, 2011), or, on the other hand, we can also speculate that the lower we go in the gymnasts ranking, the more difficult it is for the judge to identify the difficulty elements, probably due to an execution problem. We propose this point of view because the other possibility is to consider that the judges were not able to identify the difficulties performed by the gymnasts. But, as we know, in Rhythmic Gymnastics World Championship competitions only highly prepared judges can evaluate. Studies such as St. Marie et al., 2001; Plessener et al, 2005; Johansson, 2010; Dallas, 2010; Heinen, 2012; Fernandez-Villarino, 2013 and more recently (Flessas et al., 2015 have shown that the experience of the judge and her capacity to use other cognitive strategies in perceiving error may be an asset in gymnastics judging. Thus, the non-recognition of the difficulties seems to be inappropriate.

To help explain these results, we tried to understand if these differences between the DDS and FDS could be identified in the different types of difficulty elements. We found out that they exist in all difficulty elements, being significantly higher in the difficulty elements Rotations, DER and Mastery and almost residual in the difficulty elements Dance Steps for the gymnasts in the three groups of the ranking. Analysing the elements first mentioned, these results may suggest the coaches and the judges understand the CoP rules differently. In the difficulty elements with higher differences between the DDS and FDS they may have some problems in the definition of the criteria that characterises them and/or in the comprehension of the technical faults, which cancel the value of the difficulty. On the other hand, and following the same perspective, in the difficulty elements such as Dance Steps, in which there is a strong proximity to the DDS and FDS, there seems to be an almost perfect understanding of the criteria defined by the CoP.

The contribution of each type of difficulty element to the difference between DDS and FDS is higher in Rotations, DER and Mastery in the three groups of the ranking. These difficulty elements are very complex elements to perform and demand an extraordinary coordination, perfect control of the apparatus technic and a lot of practice hours (Lebre, 2011; Vitrichenko et al, 2011). Therefore, they are also the ones where the gymnast can make more technical faults which cancel the value of the difficulty, mainly the weaker gymnasts. The gymnasts with the intention of getting top scores should present routines with a high level of difficulty combined with good execution quality (Agopyan, 2014). In the case of high ranked gymnasts, this result cannot be explained by the execution scores received, because they were very high. In the case of the gymnasts ranked in groups 2 and 3 (middle and lower gymnasts), the inferior quality in the execution may justify these results, suggesting, therefore, that the coaches, do not have a real perception of the performance capacity of their gymnasts in these types of difficulty elements.

In conclusion, the results show there are significant differences between the scores proposed by the coaches and the difficulties that the judges could identify, which increase as the gymnasts go lower in the ranking. The contribution of the difficulty elements to these differences is higher in Rotations, DER and Mastery in the gymnasts placed in the three groups of the ranking. These results, can suggest that the coaches do not have a real perception of the performance capacity of their gymnasts, in these types of difficulty elements.

We suggest that future studies should examine if the difference between DDS and FDS depends on the type of apparatus.

We believe that the evaluation of the difference between DDS and FDS are variables to consider in order to help reconstruct the CoP definitions related to some types of difficulty: (i) the way coaches and judges understand the rules, (ii) the orientation of the training process to maximize the performance capacity of the gymnasts in the type of difficulty elements in which the difference of DDS and FDS is higher, and/or (iii) strategically give preference to the types of difficulties in which this difference is lower.

References

Agopyan A (2014) Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. Journal of Scientific Research 19(12): 1554-1565.

Arkaev LI, Suchilin NG (2004) How to create champions - the theory and methodology of training top-class gymnasts. Oxford: Meyer&Meyer Sport.

Ávila-Carvalho L, Leandro C, Lebre E (2011) 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), Book of abstracts of the 16th Annual Congress of the European College of Sport Science. Liverpool, UK 579-580.

Ávila-Carvalho L, Klentrou P, Palomero ML, Lebre E (2012) Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. The Open Sports Sciences Journal 5: 146-153.

Bucar PM, Cuk I, Pajek J, Kovac M, Leskosek B (2013) Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? Journal of Human Kinetics 37: 173-181.

Caburrasi EF, Santana MV (2003) Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002. Granada2003 Available from: http://www.efdeportes.com/efd65/grd.htm.

Čuk I, Fink H, Leskošek B (2012) Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. Science Gymnastics Journal 4: 73 - 82.

Dallas G, Kirialanis P (2010) Judges' evaluation of routines in men's artistic gymnastics. Science Gymnastics Journal 2: 49-58

Ferreirinha J, Carvalho J, Côrte-Real C, Silva A (2011) Evolução do Valor Real de Dificuldade dos Exercícios de Paralelas Assimétricas de Ginástas de Elite nos Ultimos Ciclos Olímpicos. In FGP (Ed.), Da Prática à Ciência. Artigos do 2º e 3º Congresso de FGP, Lisboa 71-78.

Fernandez-Villarino M, Bobo-Arce M, Sierra-Palmeiro E (2013) Practical Skills of Rhythmic Gymnastics Judges. Journal of Human Kinetics 39: 243-249.

FIG-International Gymnastics Federation (2012) Code of Points for Rhythmic Gymnastics Competitions. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

FIG-International Gymnastics Federation (2013) Gymnastics Results. Available at: http/www.gymnasticsresults.com

Flessas K, Mylonas G, et al. (2015) Judging the Judges' Performance in Rhythmic Gymnastics. Medicine & Science in Sports & Exercise. American College of Sports Medicine.

Johansson A (2010) Deciding who is the best. Validity issues in selections and judgements in elite sport. Academic dissertation, Faculty of Social Sciences, Umeå University, Sweden.

Heinen T, Vinken P, Velentzas K (2012) Judging Performance In Gymnastics: A Matter Of Motor Or Visual Experience? Science of Gymnastics Journal 4: 63-72.

Hökelmann A, Breitkreutz T, Liviotti G (2012) Changes in performance structure during group competitions in rhythmic gymnastics. Edited by Prof. Derek M. Peters & Dr. Peter G. O'Donoghue, Book of abstracts of the World Congress of Performance Analysis of Sport IX University of Worcester, UK 99.

Lebre E (2011) Technical principles for the new framework. Crossroads to the Future. International Federation of Gymnastics Scientific [Press release].

Liviotti G, Hökelmann A (2012) Which Quantifiable Performance Parameter(s) Determined the Medals Winners at the World Championship 2011 In Rhythmic Gymnastics – Group Competition?. In Rsupesy &T.(Ed.), Book of abstracts of the VI International Scientific Conference of Students and Young Scientists, Modern University Sport Science, Moscow 43.

Massidda M, Calò MC (2012) Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, Journal of Sports Sciences (30)13: 1415-1420.

Pelin RA (2013) Studies Regarding The Rhythmic Gymnastics From The Olympic Games. Sport & Society / Sport si Societate (13): 61-61.

Plessner H (2005) Positive and negative effects of prior knowledge on referee decisions in sports. T. Betsch, Haberstroh,S (Ed.) The routines of decision making 311-324.

Ste-Marie DM, Valiquette SM, Taylor G (2001) Memory Influenced Biases in Gymnastic Judging Occur Across Different Prior Processing Conditions. Research Quarterly for Exercise and Sport 72(4): 420-426.

Trifunov T, Slobodanka D (2013) The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. Physical Culture 67(2): 120-129.

Vitrichenko N, Klentrou N, Gorbulina N, Della-Chiaie D, Fink H (2011) Groups. In FIG (Ed.), Rhythmic Gymnastics. Technical Manual. Level 3, Lousanne: FIG Academy 3-55.

Wang M, Lu M, Sun X (2013) Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. Journal of Physical Education/Tiyu Xuekan (20) 5: 117.

Apendix 4. Published paper 4: Technical content of elite rhythmic gymnastics.

TECHNICAL CONTENT OF ELITE RHYTHMIC GYMNASTICS

Catarina Leandro[1,2] , Lurdes Ávila-Carvalho[3], Elena Sierra-Palmeiro[2], Marta Bobo-Arce [2]

[1]Faculty of Psychology, Education and Sport, University Lusófona of Porto, Portugal
[2]Faculty of sport Science and Physical Education, University of Coruña, Spain
[3]Sports Faculty - CIFI2D - Centro de Investigação, Formação, Intervenção e Inovação em Desporto, Porto University, Portugal

Original article

Abstract

The primary aim of this study was to analyse all technical elements used in the Rhythmic Gymnastics Kiev World Championship 2013, and identify the Structural characteristics of the technical content of elite Rhythmic Gymnastics individual routines. The data has been collected from the difficulty forms concerning 288 individual routines. To allow the comparison between gymnasts with different levels the individual routines were clustered into 3 subgroups according to their final ranking competition. Body difficulty elements were organized, according to the composition requirements stated in the RG Code of Points (FIG, 2012). Non-parametric tests - Kruskal-Wallis, Mann-Whitney and Friedman test were applied to determine whether there were significant differences between groups. As main results we can point out that in general the rhythmic gymnasts used similar body difficulties with limited variety. The highest valued elements are Dynamic Elements with Rotation and throw (DER) and rotations and these represent 50% of the total value of the routine. Concerning the dance steps and mastery, no differences were found between the routines of gymnasts place in the three parts of the ranking. The routines had differences in the composition pattern between the gymnasts according to the final ranking of the gymnasts in following items: (i) on the number of rotations of flat foot or other part of the body, Fouetté rotations and Mixed Difficulties; (ii) on the value of jumps, rotations, DER and Mixed Difficulties.

Keywords: *body difficulty, individual routines, evaluation, rhythmic gymnastics.*

INTRODUCTION

The main reason for the success in RG competition is the capacity to perform the exercise, with high level body elements and apparatus technic, with perfect execution, in harmony with the character and rhythm of the music, respecting the principle of originality and diversity. This is a guarantee of an exciting high performance sport to watch.

The rules which guide the routines composition can also have influence in the gymnasts' performance (Massidda, 2012).

The limited variety on the choice in difficulty elements makes the routine composition boring and puts in risk its artistic value (Ávila, 2012a).

The skilful interaction between the gymnast and the apparatus and the increase difficulty elements in the routines composition are the development in RG (Lebre, 2011).

The analysis of these factors can, according to Ávila, (2012b), influence the developmental programs for the practice and the experimental designs used in the scientific research in RG. The knowledge can also contribute to preview and characterize the effort requirements allowing improvements in the gymnasts' preparation to the competition readiness (Ferreirinha, 2009).

Ferreirinha (2009) refers that to determine the training models it is important to know the characteristics competition routines for high level gymnasts including the details concerning the specificity of their components.

Is, than, fundamental to analyse the development tendencies for the sport in general and to identify specificities of each component as we propose to do with the structural characteristics of the difficulty elements including the diversity and variety in the routines.

The routines composition is not stable concerning their content because they have to be adapted to changes done in the Code of Points (CoP) every Olympic cycle.

An individual RG routine is composed by a series of body and apparatus elements linked in a specific way which we call difficulty elements (D). On the present Olympic cycle, the content of and individual RG routine should respect the specific requirements that are common to the routines of all 4 apparatus: jumps/leaps, balances, rotations, apparatus mastery, dance steps and dynamic elements with rotation and throw (DER) (FIG, 2012).

The value of each difficulty element is from 0.10 points to 1.50 points or more, which may be absolutely determinant in the final score obtained in competition. The inclusion of complex abilities in the routines is essential to have a high score (Massida, 2012).

The CoP holds a great variety of difficulty elements to be used in the routines. One important characteristic of the RG is to allow the gymnast to link it in her own way, with a stylish presentation, clever configuration, and prefect presentation (Wang, 2013). An eventual lack of variety in the body difficulty included in the routines can cause judges and audience dissatisfaction from the point of originality and variety. RG is a visually appealing sport, thus, it is very important to keep the high interest of the public (Agopyan, 2014).

The studies published concerning the content of the RG routines (Caburrasi, 2003; Bobo, 2010; Ávila, 2011; Ávila, 2012; Trifunov, 2013; Agopyan, 2014), include the analysis of the number and the level of difficulty elements but they have no information about contribution of each type of element for the final D score. Also in consequence of the constant and quickly evolution of this sport, a permanent upgrade of these kind of studies are needed (Caburrasi, 2003; Cuk, Fink & Leskošek, 2012; Massidda, 2012; Hökelmann et al., 2012; Bucar, 2013; Pelin, 2013).

Within this context, the main goal of the present study is to identify the difficulty elements included in the routines that contribute the most to the success in competition and to analyse the diversity of the body movements included in the difficulty elements.

The present study can have an important contribution for the coaches mainly to: supporting the coaching process, defining performance profiles for individual gymnasts, ranking performances, creating data bases in order to identify the most influencing performance indicators and the tendencies in the development of RG (Liviotti, 2012).

METHODS

288 difficulty forms concerning individual routines were analysed. The routines were performed by gymnasts from 45 different countries competing at Rhythmic Gymnastics World Championship in Kiev, Ukraine in 2013. This study was done with the permission of the International Gymnastics Federation (FIG).

The official Difficulty forms, submitted prior to the competition, included the routine compositions recorded using the RG CoP symbols. All difficulty elements reported in the difficulty forms were analysed. The analyse was done considering the all sample, and the sample clustered into 3 subgroups according to gymnasts final ranking as follows: the first part of the ranking - the top 24 gymnasts, the second part of the ranking - 24 middle gymnasts and third part of the ranking – the 24 lower placed gymnasts on the ranking, to allow the comparison the technical elements within gymnasts of different levels.

The analysis was conducted by two international RG judges. The intraclass correlation coefficient (ICC) in test-retest method (intra-examiner) was 0.99. The ICC between the observers (inter-examiner) was 0.98.

The data were analyzed using the Statistical Package for Social Sciences – version 20.0 (SPSS 20.0, Chicago, USA) and Microsoft Office Excel 2007. The level of significance was set at $\alpha = 0.05$ (confidence interval of 95%). Descriptive statistics were calculated using the mean values as a measure of central tendency, standard deviation (SD) as a measure of dispersion, and minimum and maximum as measures of data range. After checking the abnormalities in the data distribution ($p<0.05$) using the Kolmogorov-Smirnov normality test, we resorted to non-parametric test (Kruskal-Wallis, Mann-Whitney and Friedman test) to determine whether there were significant differences between the three subgroups in the Rhythmic Gymnastics World Championship ranking.

A multiple regression was used to analyze the influence of each difficulty element in the gymnasts' final difficulty score.

RESULTS

The difficulty elements reported in the individual routines were grouped by technical categories: balances, jumps, rotations, masteries, dance steps, and DER, mixed difficulties (MixDif) and criteria associated to difficulty (waves and pre-acrobatics). The results for each category are presented both quantitatively (number of occurrences) and qualitatively (technical value and type) in Figure 1. From Figure 1 we can highlight the number of the mastery (4.0±2.80) and the value of the rotations with 29% of the total value of the routine (2.7±0.83 points). When we observe the three difficulty groups that are based on the body movements (jumps, balances and rotations) we can see that the rotations have the higher number (3.3±0.61) and the balances the lower number (2.4±1.00). Concerning the rotations, the gymnasts included preferably those with 0.30 points values in their routines. Between them it is possible to highlight the "pivot attitude" (0.52±0.50), the "pivot free leg in ring in back with help" (0.42±0.50) and the "rotation penché" (0.76±0.43). The most used jumps were those with 0.5 points value, mainly the "jeté with turn" (0.82±0.80) and the "jeté with a turn with back bend" (0.45±0.53); The balances with base value 0.5 points were the most performed by the gymnasts, mainly the balance "side scale with split, without help" (0.44±0.49) and balance "back scale leg high up" (0.40±0.49). The most used MixDif were the link of the balance "front scale with back split" and "ring without help" (0.15±0.52). For DER, the most used criteria to raise the value were: "change of level", "change of body rotation axis", "throw/catch outside of visual

control" and "throw/catch without the help of the hands".

Analysing the sample according to final ranking of the gymnasts, significant differences were found on the number of balances, MixDif, rotations on the flat foot or other part of the body, and "fouetté" rotations (Table 1). No other significant differences in the number of technical difficulties were found according to the final ranking of the gymnasts (Figure 2).

The number of balances was significantly higher in the gymnasts of the 3rd part of the ranking and the MixDif significantly higher in the gymnast of the 1st part of the ranking. The number of rotations on the flat foot or other part of the body is higher in the 1st part of the ranking and decreases significantly in the 2nd and 3rd parts. The number of "fouetté" rotations is significantly higher in the gymnast of the 2nd part of the ranking (Table 1).

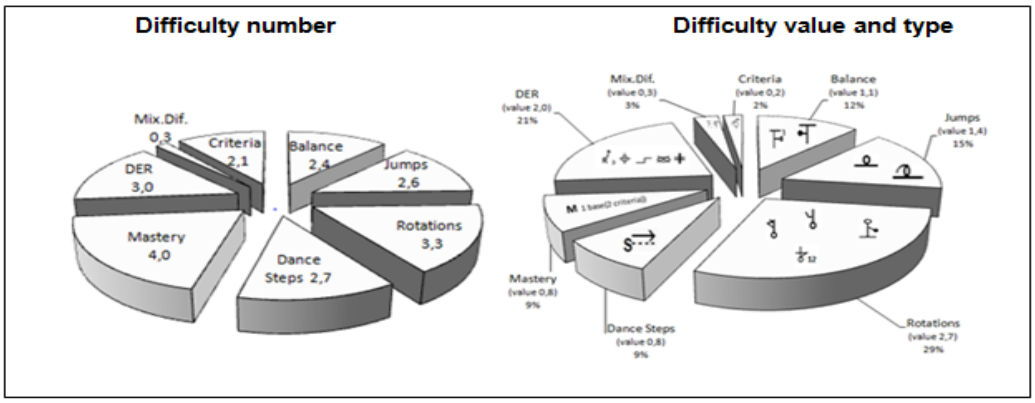


Figure1. Difficulty elements presented in the Rhythmic Gymnastics individual routines in the 2013 World Championships clustered according to number, value and type.

Table 1
Number of balances, MixDif and Rotations in the Rhythmic Gymnastics individual routines clustered according to gymnasts' final ranking in the 2013 World Championships.

	1st part of the ranking (n=96)		2nd part of the ranking (n=96)		3rd part of the ranking (n=96)		Kruskal-wallis test	**Pairwise Comparisons**
Difficulty number	Mean+sd	Min-Max	Mean+sd	Min-Max	Mean+sd	Min-Max	P=	
Balance	2.16±1.08	0-4	2.48±0.91	1-4	2.66±0.93	0-4	0.002*	rk1-rk3 0.001
Mix. Dif.	0.75±0.97	0-2	0.56±0.90	0-2	0.38±0.78	0-2	0.016*	rk3-rk1 0.012
Rot.flat foot..	1.03±0.49	0-2	1.00±0.50	0-2	0.68±0.53	0-2	0.000*	rk3-rk1 0.000 rk3-rk2 0.000
Rot. "Fouette"	0.28±0.49	0-2	0.65±0.69	0-2	0.58±0.57	0-2	0.000*	rk1-rk3 0.001 rk1-rk2 0.000

*p<0,05

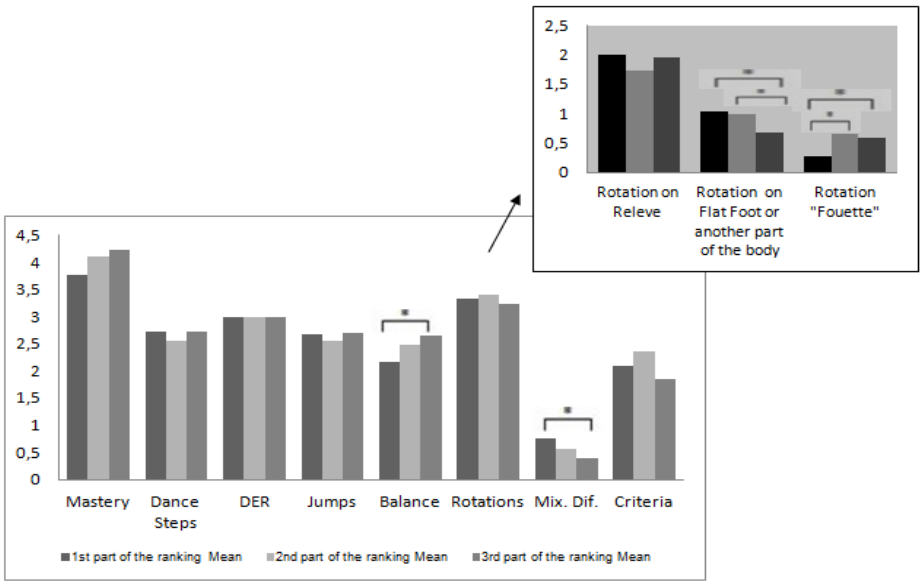


Figure 2. Average number of difficulty elements presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking.(*p<0.05).

Table 2
Technical value of DER, Jumps, MixDif and Rotations presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking.

	1st part of the ranking (n=96)		2nd part of the ranking (n=96)		3rd part of the ranking (n=96)		Kruskal-wallis test	Pairwise Comparisons
Difficulty value	Mean+sd	Min-Max	Mean+sd	Min-Max	Mean+sd	Min-Max	P=	
DER	2.05±0.30	1.3-2.7	2.01±0.27	1.4-2.8	1.83±0.35	1.0-2.8	0.000*	rk3-rk2 0.001 rk3-rk1 0.000
Jumps	1.54±0.44	0.7-2.8	1.4±0.45	0.7-2.7	1.25±0.33	0.5-2.3	0.000*	rk3-rk1 0.000
Mix. Dif.	0.37±0.49	0.0-1.7	0.25±0.41	0.0-1.0	0.16±0.34	0.0-1.0	0.005*	rk3-rk1 0.003
Rotations	3.00±0.88	1.2-4.7	2.89±0.66	1.2-4.3	2.22±0.71	0.7-3.9	0.000*	rk3-rk2 0.000 rk3-rk1 0.000
Rot. Releve	1.85±0.97	0.0-4.0	1.56±0.88	0.0-3.3	1.37±0.67	0.0-3.1	0.001*	rk3-rk1 0.001
Rot.flat foot..	0.89±0.44	0.0-2.3	0.80±0.37	0.0-1.7	0.49±0.40	0.0-1.4	0.000*	rk3-rk2 0.000 rk3-rk1 0.000
Rot.Fouette	0.25±0.47	0.0-2.0	0.53±0.59	0.0-2.0	0.36±0.38	0.0-1.2	0.001*	rk1-rk2 0.001

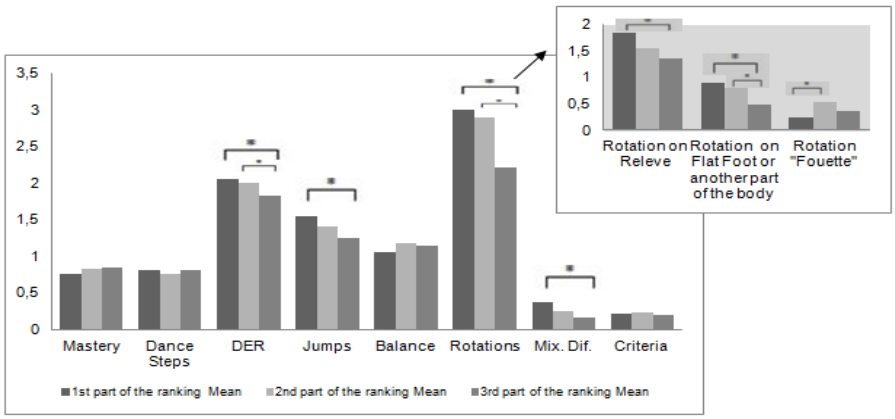


Figure 3. Technical value of the difficulty elements presented in the Rhythmic Gymnastics individual routines clustered according to the 2013 World Championships final ranking. (*p<0.05)

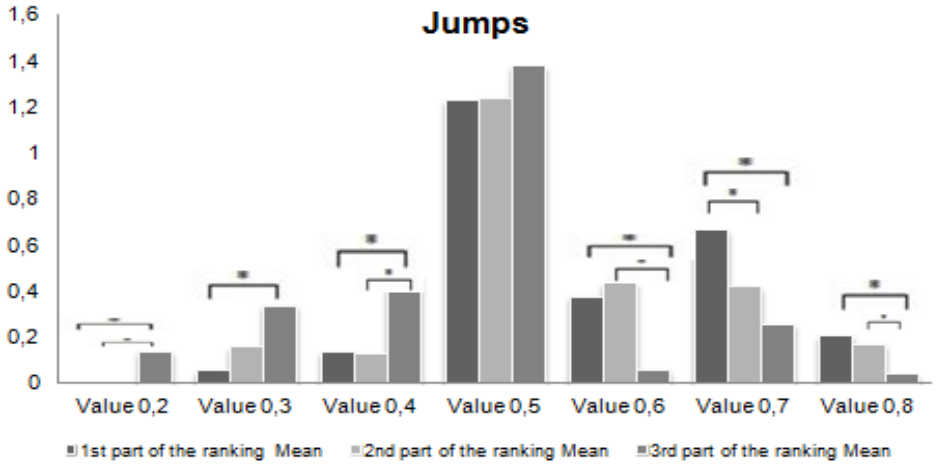


Figure 4. Number of Jump difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships (Kruskal-wallis test * p<0,05).

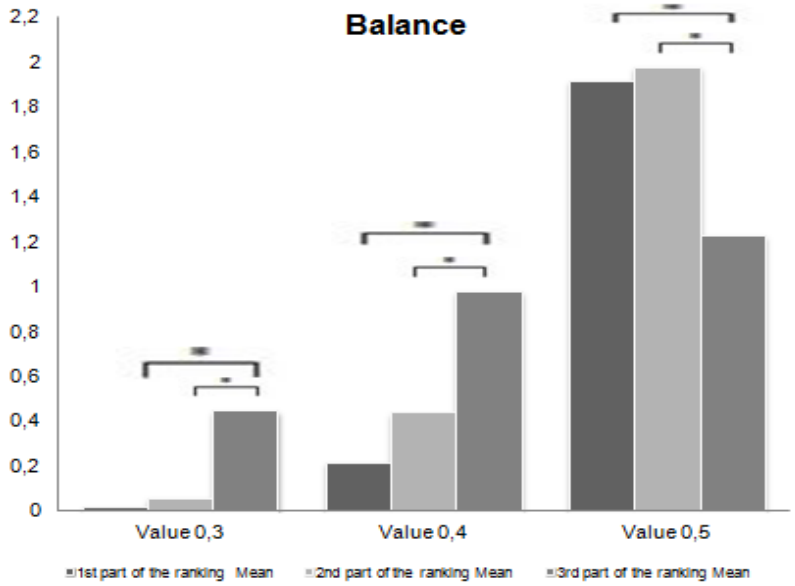


Figure 5. Number of balance difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships (Kruskal-wallis test * p<0,05).

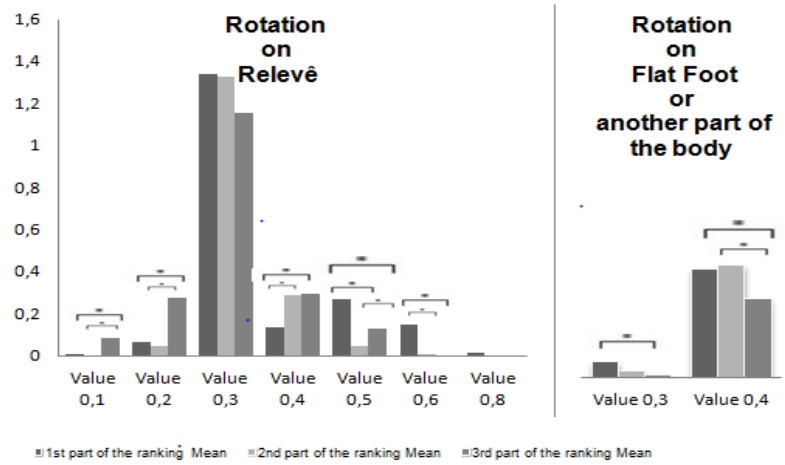


Figure 6. Number of Rotations difficulties (different values) presented in the Rhythmic Gymnastics individual routines clustered according to the gymnasts' final ranking in the 2013 World Championships. (Kruskal-wallis test * p<0,05).

The value of the DER, rotations, jumps and MixDif is higher in the gymnast placed in the 1 st part of the ranking and decreases significantly in the 2nd and 3rd parts. For mastery, dance steps, balances and criteria associated to difficulty (waves and pre-acrobatics) there were no statistically significant differences regarding the technical value, and the gymnasts' final ranking (Table 2). Concerning the value of the rotations we can highlight that the fouetté rotations had a significant higher value in the gymnasts placed on the 2nd part of the ranking when compared to the gymnast in the 1 st part. The rotations of flat foot or another part of the body registered a higher value in the 1st part of the ranking (Figure 3).

Jumps

Analysing the value of the jumps included in the routines according to the final ranking of the gymnasts, we found significant differences for all jumps except the jumps with 0.5 points value. The routines of the gymnasts placed in 1st part of the ranking had a higher number of jumps 0.7 and 0.8 points value. The jumps of value 0.2, 0.3 and 0.4 points were the preferred of gymnasts placed in the 3rd part of the ranking. The jumps with 0.6 points value are performed preferably by the gymnasts on the 2nd part of the ranking. The jumps with 0.5 points value jumps were the preferred of all gymnasts independently of their place on the final ranking. There were not significant differences for the gymnasts ranking regarding the jumps of 0.5 value jumps (Figure 4).

Balances

We found significant differences in the number of balance difficulties when we compare the routines performed by the gymnasts of different parts of the final ranking. The gymnasts ranked in the 3rd part of the ranking had a higher number of balances with 0.30 and 0.40 points value in their routines. On the other hand the gymnasts ranked on the 1st and 2nd parts preferred to include in their routines balances of 0.50 points (Figure 5).

Rotations

There were significant differences in the rotations included in the routines in all parts of the final ranking excepted for the rotation of 0.5 points value. The routines of the gymnasts ranked in the 3rd part had a higher number of rotations on "relevé" of 0.1, 0.2 and 0.4 points value. On opposite, the gymnasts placed in the 1st and 2nd parts preferred to include 0.5 and 0.6 points value rotations on "relevé" in their routines. The rotations on relevé with 0.3 points value were the most performed by all gymnasts independently of their position in the final ranking (Figure 6). Concerning the rotations on flat foot or another part of the body, it was clear that the gymnasts placed in the 1st and 2nd parts of the ranking preferred to include this type of rotation with 0.4 points value in their routines.

DISCUSSION

This study provides a quantitative and qualitative analysis of the difficulty elements used in the individual routines of the 2013 RG World Championships.

The 288 individual routines studied were clustered into three subgroups according to the gymnasts' final ranking in the World Championships. We discussed the results (number, value and type) in 3 dimensions: (1) global analysis of the composition of the routines; (2) analysis by group of difficulty elements; (3) ranking of the gymnasts.

In a global point of view the routines hold an average value of 9.30 points, very close to the maximum possible score of 10 points. Despite the World Championships being one of the most important competition in the calendar, this result may lead to a false analysis, as it could mean such a high a plateau of international excellence which in reality only occurs amongst gymnasts at the top of the ranking. Ávila, (2011) studied the difference between the departure score (presented by

the coach in the difficulty form) and the final score obtained by the gymnast and concluded that the majority of the gymnasts reach very significant differences, of 2 or more points between these two scores.

We also highlight the fact that the rotations and the DER, together represented 50% of the total value of the composition. This result showed an important change in the global content of the routines in this Olympic Cycle. Studies such as Caburrasi (2003) and Ávila (2011) showed that in the previous Olympic cycles the highest contribution in the routines value came from the Jumps. The increase in the rotations and DER values happens because in the present Olympic cycle it is possible to add some criteria to these difficulties that allow the gymnast to increase its value and degree of complexity (FIG, 2012). These results can be analysed in two different perspectives. On the one hand, it represents an upgrade of the execution quality, but on the other hand, it means that an extreme importance is given to 2 types of difficulty elements leading to an under estimation of the other groups. We also remarked the lack of variety and diversity in the elements chosen that has been repeatedly mentioned in previous studies concerning individual routines (Bobo, 2003; Agopya, 2014) and group routines (Ávila, 2011b; Ávila 2012; Ávila, 2012b). Therefore it is possible to conclude that the RG routines present a consistent pattern in the usage of the difficulty elements.

The type of difficulty elements used in the routines is similar, with some difficulty elements being repeated several times in the routines. This means that the routines' composition is not defined by being unique, with diversity and creativity, characteristics that are necessary for the enrichment of the routines composition (Balcells, 2009; Leandro, 2015) and reflect the spectacular of the choreography (Pelin, 2013).

The analysis according to the type of difficulty showed us similar results as Agopyan (2014) for routines performed on the last Olympic cycle: the rotation difficulties (mainly the "relevé" rotations) were the preferred of the gymnasts and the balance difficulties the less used. The rotation difficulties are very complex elements to perform (Lebre, 2011; Vitrichenko et al, 2011), but they are also those where the gymnast can get more points, once the CoP (FIG,

2012) allows to add the base value of the difficulty for each rotation performed. The lower number of balances in the routines is, probably, due to the fact that the gymnasts spend considerable time of the routine to perform them because they are static difficulty elements (Gateva, 2015) and they have low values: 0.50 points is the maximum possible value for a balance, according to the CoP (FIG,

2012). These are the main reasons for the preference of the gymnasts to include more difficulties in rotation and less in balance in their routines. The routines only last for maximum 90 seconds and they have to optimize the time available to get the maximum of points allowed (10 points). The gymnasts, with the intention of getting top scores should present routines with a high level of difficulty combined with good execution quality (Agopyan, 2014).

The mastery and dance steps have comparatively lower possible values than the jumps, rotations and balances. These groups have an inferior degree of execution complexity, they are less valued in the CoP (FIG, 2015). To promote the inclusion of these types of elements in RG routines, and therefore have more interesting choreographies their value should be increased (Livotti, 2012; Leandro, 2015). One of basic requirements of RG is that the gymnast should show an optimal use of the body together with the apparatus handling. In this way, to raise the difficulty departure score the gymnast must increase of both body and apparatus difficulty level included in the routine (Agopyan, 2014).

The analysis of the results according to the gymnasts' final ranking showed that the higher placed gymnasts chose preferentially elements with a higher complexity (MixDif, rotations on flat foot or other part of the

body and "fouetté" rotations) and the lower placed gymnasts chose elements with lower complexity (balances) as described also by Gateva (2015).

Regarding the difficulty elements value, the jumps were the elements with higher value for the gymnasts in the first and second parts of the ranking. With the exception of the jumps of 0.5 points value, the gymnasts higher placed (1st and 2nd part of the ranking) include preferably the jumps with higher value and the gymnast placed on 3rd part preferred the jumps of 0.3 and 0.4 points value, which confirms the expectable. According to Bobo (1998) and Bobo, (2003), as a norm the best gymnasts hold physical and artistic capacities that allow them to perform more and higher level elements with high execution complexity. The rotations, DER and MixDif had higher values in the routines of the gymnasts placed in 1st part of the ranking and decreased in the routines of the gymnasts placed in the second and third parts. The complexity of this type of difficulties is very high and demands an extraordinary coordination, a perfect control of the apparatus technic and a lot of practice hours, (Lebre, 2011; Vitrichenko et al, 2011), which justifies that they are preferably used by the gymnasts highly ranking.

CONCLUSIONS

The rhythmic gymnasts who competed at the 2013 World Championships used in their routines very similar difficulties elements with limited variety. The more used difficulties were the rotation "attitude", rotation with "free leg in ring in back with help", "rotation in penché"; balance "side scale with split, without help" and balance "back scale leg high up"; jump "jeté with turn" and "jeté with a turn with back bend". The highest valued elements are DER and rotations and these represent 50% of the total value of the routine. These groups showed an important contribution to the final D score. The balances were the less used difficulty group.

The routines had differences in the composition pattern between the gymnasts according to the their final ranking in the following items: (i) the number of rotations of flat foot or other part of the body, "fouetté" rotations and MixDif; (ii) the value of jumps, rotations, DER and MixDif. Concerning the dance steps and mastery, no differences were found between the routines of gymnasts place in the three parts of the ranking.

This study provides updated information about the individual routines content in rhythmic gymnastics, to be considered: (i) to the possible modifications of the present Code of Points, in particular for the definition of the composition requirements in order to have higher level of variety and diversity in the routines, and (ii) to the training process to achieve the high performance level in the individual gymnasts.

REFERENCES

Agopyan, A. (2014). Analysis of Body Movement Difficulties of Individual Elite Rhythmic Gymnasts at London 2012 Olympic Games Finals. *Journal of Scientific Research, 19*(12), 1554-1565.

Arkaev, L.I. & Suchilin, N.G.(2004). *How to create champions - the theory and methodology of training top-class gymnasts*. Oxford: Meyer&Meyer Sport.

Ávila-Carvalho, L., Klentrou, P., Palomero, M. d. L., & Lebre, E. (2012) Analysis of the Technical Content of Elite Rhythmic Gymnastics Group Routines. *The Open Sports Sciences Journal,* 5, 146-153.

Ávila-Carvalho, L., Klentrou, P., & Lebre, E. (2012). Handling, Throws, Catches and Collaborations in Elite Group Rhythmic Gymnastics. *Science of Gymnastics Journal, 4*(3), *37-47*.

Ávila-Carvalho, L., Leandro, C., & Lebre, E. (2011). 2009 Portimão Rhythmic Gymnastics World Cup. Scores analysis. In N. T. Cable & K. George (Eds.), *Book of abstracts of the 16th Annual Congress of the European College of Sport Science*. Liverpool, UK. 579-580.

Ávila-Carvalho, L., Palomero, M. d. L., & Lebre, E. (2011b). Estudio del valor artístico de los ejercicios de conjunto de Gimnasia Rítmica de la Copa del Mundo de Portimão 2007 y 2008. *Apunts. Educación Física y Deportes, 1.er trimestre* 103, 68-75.

Ávila-Carvalho, L., Palomero, M. d. L., & Lebre, E. (2009c). Difficulty score in Group Rhythmic Gymnastics. Portimão 2007/2008 World Cup Series. *Palestrica Mileniului III. Civilizatie si sport, Anul X, 3*(37), 261-267.

Balcells, M., Martín, C., & Anguera, M. (2009). *Instrumentos de observación ad hoc para el análisis de las acciones motrices en Danza Contemporánea, Expresión Corporal y Danza Contact-Improvisatio*. Apunts educación física y deportes.Ciencias aplicadas a la actividad física y el deporte, 14-23.

Bobo M, Sierra E.(2003). *Estudio de las repercusiones de los cambios de código de puntuación en la composición de los ejercicios de gimnasia rítmica en la técnica corporal.* Available from: http://www.cienciadeporte.com/congreso/04%20val/pdf/p3.pdf.

Bobo M, Sierra E. (1998). *Una nueva propuesta de dificultades corporales en gimnasia rítmica deportiva*. Libro de resúmenes del VI Congreso de Educación Física e Ciencias do Deporte dos Países de Lingua Portuguesa Deporte e Humanismo en clave de Futuro.

Bucar, P. M., Cuk, I., Pajek, J., Kovac, M., & Leskosek, B. (2013). Is the Quality of Judging in Women Artistic Gymnastics Equvalent at Major Competitions of Different Levels? *Journal of Human Kinetics*, 37, 173-181.

Caburrasi, E.F., Santana, M.V. (2003). *Análisis de las dificultades corporales en los Campeonatos Europeos de Gimnasia Rítmica Deportiva Granada 2002*. Available from: http://www.efdeportes.com/efd65/grd.htm.

Čuk, I., Fink, H., & Leskošek, B. (2012). Modeling The final score in Artistic Gymnastics by different weights of difficulty and execution. *Science of Gymnastics Journal*, 4, 73 – 82.

FIG. (2012). *Code of Points for Rhythmic Gymnastics Competitions*. Available at: http://www.fig-gymnastics.com/site/page/view?id=472

Ferreirinha, J., Carvalho, J., Côrte-Real, C., & Silva, A. (2011). Evolução do Valor Real de Dificuldade dos Exercícios de Paralelas Assimétricas de Ginástas de Elite nos Ultimos Ciclos Olímpicos. *In FGP* (Ed.), Da Prática à Ciência. Artigos do 2º e 3º Congresso de FGP (pp. 71-78). Lisboa.

Gateva, M., Gospodarski, N., Treneva, V., Avramov, D., Ivanov, N., & Andonov, K. (2015). Comparison Between The Static Balance Of Practitioners From Different Sports and Non-Athletes. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science*. p. 569. Malmo, Sweden.

Hökelmann, A., Breitkreutz, T., & Liviotti, G. (2012). *Changes in performance structure during group competitions in rhythmic gymnastics*. Edited by Prof. Derek M. Peters & Dr. Peter G. O'Donoghue, *Book of abstracts* of the World Congress of Performance Analysis of Sport IX University of Worcester, UK. p99.

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). Accuracy in Judgment the Difficulty Score in Elite Rhythmic Gymnastics Individual Routines. *Science of Gymnastics Journal, 7*(3):81-93.

Leandro, C., Ávila-Carvalho, L., Sierra-Palmeiro, E., & Bobo-Arce, M. (2015). What Do Rhythmic Gymnastics Judges Think About Their Code Of Points?. Edited by Radmann, A., Hedenborg, S., Tsolakidis, E. *Book of abstracts of the 20th Annual Congress of the European College of Sport Science* (pp. 569). Malmo, Sweden.

Lebre, E. (2011). Technical principles for the new framework. Crossroads to the Future. International Federation of Gymnastics Scientific [Press release].

Liviotti, G. & Hökelmann, A. (2012). *Which Quantifiable Performance Parameter(s) Determined the Medals Winners at the World Championship 2011 In Rhythmic Gymnastics – Group Competition?*. In Rsupesy &T.(Ed.), *Book of abstracts* of the VI International Scientific Conference of Students and Young Scientists "Modern University Sport Science" . Moscow, p43.

Liu, X.X., & Kuang, L. (2001) Review of evolvement course of international evaluation rules in rhythmic gymnastics and its effects on technique development. *Journal of Beijing Sport University, 3*(24), 412-415.

Massidda, M. & Calò, M.C. (2012). Performance scores and standings during the 43rd Artistic Gymnastics World Championships, 2011, *Journal of Sports Sciences, 13* (30), 1415-1420.

Pelin, R. A.(2013) Studies Regarding The Rhythmic Gymnastics From The Olympic Games. *Sport & Society / Sport si Societate*, 13, 61.

Sands, W.A. ,Caine, D.J., & Bornes, J. (2003). Scientific aspects of women´s gymnastics .(vol.45). Basel: S. Karger A.G. Trifunov, T., & Slobodanka, D. (2013). The structure of difficulties in the routines of the best world and serbian rhythmic gymnasts. *Physical Culture, 67*(2), 120-129.

Vitrichenko, N., Klentrou, N., Gorbulina, N., Della Chiaie, D. & Fink, H. (2011). Groups. In FIG (Ed.), *Rhythmic Gymnastics. Technical Manual. Level 3.* 3-55. Lousanne: FIG Academy.

Wang,M., Lu,M., & Sun, X. (2013). Structural characteristics of the rhythmic gymnastic difficulty system examined from the perspective of the new rules. *Journal of Physical Education / Tiyu Xuekan*, *5*(20), 117.

Corresponding author:

Catarina Leandro
Faculty of Psychology, Education and Sport
University Lusófona of Porto
Portugal
Phone/mobile: +351917640998
E-mail: catarinaleandro@sapo.pt